

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 18, 2000, 20:18:44 ; Search time 360.16 Seconds
(without alignments)
-3498.721 Million cell updates/sec

Title: US-09-092-296-7_COPY_5_419
 perfect score: 415
 Sequence: 1 ACCGGGACCTCAGTGCTCC.....AAATTCGGTATGCTGAATT 415

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl:★

Word size : 0

Number of hits that pass the threshold : 1642386

```
1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_p11:*
```

50: gb_p13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

SUMMARIES									
Result	Query No.	Score	Query			DB	ID	Description	
			Match	Length					
1	21	53.3	47323	11	AC005937	AC005937	Homio sapi		
2	219.4	52.9	192650	10	AB073038	AB073038	Homio sapi		
3	219.4	52.9	200000	10	AF000511	AF000511	Homio sapi		
C	42.2	10.2	72718	5	I66494	I66494	Escherichia coli		
	37.2	9.0	185548	45	AC0012467	AC0012467	Homio sapi		
	36.4	8.8	18276	11	AC004148	AC004148	Homio sapi		
5	36.2	8.7	65870	12	AC005807	AC005807	Homio sapi		
6	35.4	8.5	207984	4	AC006339	AC006339	Homio sapi		
7	35	8.4	47092	44	AC0013564	AC0013564	Homio sapi		
8	34.8	8.4	148231	42	AC0011995	AC0011995	Homio sapi		
9	34.6	8.3	65673	42	AC0012184	AC0012184	Homio sapi		
C	34.2	8.3	77322	10	H5D1144C9	H5D1144C9	Human DNA		
C	34.2	8.2	3626	12	D86631	D86631	Mus sp. DANA		
13	34	8.2	36676	11	H537858816	H537858816	Human DNA		
15	33.8	8.1	62181	44	AC016298	AC016298	Homio sapi		
16	33.6	8.1	128379	11	H53773A18	H53773A18	Human DNA		
17	33.6	8.1	26065	43	AC004817	AC004817	Drosophilli		
C	33.4	8.0	112309	11	AC0030325	AC0030325	Human Chr		
19	33.4	8.0	185035	11	AC004770	AC004770	Homio sapi		
20	33.4	8.0	196080	40	AC004228	AC004228	Homio sapi		
C	33.2	8.0	142092	41	AF139813	AF139813	Homio sapi		
21	33.2	8.0	167000	1	AP000059	AP000059	Aeropyrium		
22	33.2	8.0	147762	10	H5198C21	H5198C21	Human DNA		
23	33.2	8.0	108187	43	AC009581	AC009581	Homio sapi		
24	33.2	8.0	116218	11	AC005090	AC005090	Homio sapi		
C	33	8.0	124700	35	AC005558	AC005558	Drosophilli		
C	32.8	7.8	30432	10	AB014077	AB014077	Homio sapi		
C	32.8	7.8	130000	10	AF0000513	AF0000513	Homio sapi		
C	32.8	7.8	159500	11	AC004209	AC004209	Homio sapi		
C	32.8	7.8	159500	12	CC014270	CC014270	Mus muscu		
C	32.8	7.8	159500	34	AC001111	AC001111	Caenorhabdi		
C	32.6	7.8	15280	9	H5197N316	H5197N316	Homio sapi		
C	32.6	7.8	172048	11	NS97K21	NS97K21	Human DNA		
C	32.6	7.8	156197	11	AB001291	AB001291	Human DNA		
C	32.4	7.8	2458	8	AF053391	AF053391	Trichostema		
C	32.4	7.8	184427	16	EX0120824	EX0120824	Equine herpes		
C	32.2	7.8	13882	2	AE001157	AE001157	Porcine herpes		
C	32.2	7.8	3122	3	PIGCHAT	PIGCHAT	Pig choline		
C	32.2	7.8	1277	7	EMERG589	EMERG589	Human DNA		
C	41	32.2	7.8	123203	40	AC0010336	AC0010336	Homio sapi	
C	42	32.2	7.8	167228	11	AC0005552	AC0005552	Homio sapi	
C	43	32.2	7.7	37036	1	MS9151	MS9151	Mycobacte	
C	44	32	7.7	32514	1	MTCT130	MTCT130	Mycobacte	
45	32	7.7	32514	1	MTCT130	MTCT130	Mycobacte		

ALIGNMENTS

RESULT 1

			DNA	PRI	05-NOV-1998
AC005937	AC005937	47323 bp			
LOCUS	HOMO sapiens clone UWG:370M23_002 from 6p21.				
DEFINITION					
ACCESSION	AC005937				
VERSION	AC005937.1 GI:3845393				
KEYWORDS	HTG.				
SOURCE	human				
ORGANISM	Eukaryota; Metazoa;				
	Eumetazoa; Chordata; Craniata; Vertebrata; Mammalia,				
	Eutheria; Primates; Catarrhini; Hominidae; Homo				
REFERENCE	1 (bases 1 to 47323)				

AUTHORS Janer M., Guillaudoux, T., Vu, O., Kutyavain, T., Harter, H. and Geraghty, P.E.

TITLE Large scale sequence analysis of the human MHC class I region

JOURNAL Unpublished (1998)

REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024

REFERENCE 2 (bases 1 to 47323)

AUTHORS Geraghty, D.E. and Olson, M.V.

TITLE Direct Submission

JOURNAL Submitted (05-NOV-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhrc.org)

COMMENT Overlapping Sequences:
5': UWGC:370M23.013 (Genbank Accession: AC005530)
3': UWGC:y67c112 (Genbank Accession: AC004211)

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 75.5%
DS or two chemistry coverage: 98.5%
Single stranded regions:

Sequence Validation:
This sequence has been validated by Multiple Complete Digest (MCD) and has been validated by Multiple Complete Digest (MCD) and has been validated by Multiple Complete Digest (MCD). Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

	Map	Seq	HindIII	Map	Seq	Map	Seq
	BglII						NsiI
	1069.11	1050.00	889.55	866.00	30541.40	30653.00	
	20320.67	20855.00	1050.18	1015.00	3279.08	3231.00	
	2171.50	2147.00	7268.78	7196.00			
	2560.20	2531.00	10085.80	9992.00			
	4335.42	4269.00	11212.78	11131.00			
	2698.62	2628.00					
	1927.50	1887.00					
	3130.46	3090.00					
	2166.69	2129.00					
	2044.67	2005.00					

FEATURES Location/Qualifiers

source 1. 47323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21"
/subclone="UWGC-370M23.002"

/clone.lib="Research Genetics BAC Library"
3647..3932
/rpt_family="Alu"
complement(4999..5277)
/rpt_family="Alu"
6285..6572
/rpt_family="Alu"
complement(6972..7050)
/rpt_family="MLT1"
7286..7584
/rpt_family="Alu"
complement(8164..8609)
/rpt_family="Alu"
complement(21287..21895)
/rpt_family="Alu"
22715..22957
/rpt_family="Alu"
25510..25802
/rpt_family="Alu"
27835..28010
/rpt_family="MER20"
31295..31594
/rpt_family="Alu"
33515..33767
/rpt_family="Alu"
34233..34290
/rpt_family="MIR"
37372..37848
/rpt_family="Alu"
38526..38700
/rpt_family="MER3"
39585..40010
/rpt_family="Alu"
40016..40156
/rpt_family="Alu"
43194..43372
/rpt_family="MERS"
43375
/note="clonal variation with 3' overlapping clone"
4419
/note="clonal variation with 3' overlapping clone"
44451
/note="clonal variation with 3' overlapping clone"
44537
/note="clonal variation with 3' overlapping clone"
44814
/note="clonal variation with 3' overlapping clone"
44965
/note="clonal variation with 3' overlapping clone"
45760
/note="clonal variation with 3' overlapping clone"
45900
/note="clonal variation with 3' overlapping clone"
46851
/note="clonal variation with 3' overlapping clone"
46859
/note="clonal variation with 3' overlapping clone"
47032
/note="clonal variation with 3' overlapping clone"
47240..47256
/note="clonal variation with 3' overlapping clone"
insertion of 17bp repeat
BASE COUNT 11556 a 11489 c 12284 g 11994 t
ORIGIN

Query Match 53.3%; Score 221; DB 11; Length 47323;
Best Local Similarity 100.0%; Pred. NO. 1.5e-55;
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 AGCTTGGCTCCCTCCATGCTCCCTGAGGACGCGTCCGACCTCCACCATGCGAGAT 254
DB 35687 AGCTTGGCTCCCTCCATGCTCCCTGAGGACGCGTCCGACCTCCACCATGCGAGAT 35746


```

/db_xref="GBD:575281"
complement(164525..164655)
/note="SHGC-3064:The location is between each flanking
site of PCR primers."
/db_xref="GBD:1234116"
complement(169209..169574)
standard_name="D6S952"
/note="UT5233:The location is between each flanking site
of PCR primers."
/db_xref="GBD:313481"
BASE COUNT      52605 a 47531 c 43666 g 50498 t
ORIGIN

Query Match      52.9%; Score 219.4; DB 10; Length 200000;
Best Local Similarity 99.5%; Pred. No. 5e-55;
Matches 220; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 195 AGCTCTGCTCCTCCATCTCCCTTCAGGAGCAGCGTCAACCTCCACACATGCACAGAT 254
Db 110528 AGCTCTGCTCCTCCTCCATCTCCCTTCAGGAGCAGCGTCAACCTCCACACATGCACAGAT 110587

QY 255 CTCACACACATGTTGTCTCTCAACACATGACACGCATTGAAGCCCTGTCTCTTGCGCC 314
Db 110588 CTCACACACATGTTGTCTCTCAACACATGACACGCATTGAAGCCCTGTCTCTTGCGCC 110647

QY 315 GGGCTTTTGGCGGGGATGACGAGGAGCGGCCGACCTGTCTTCAGACGCGCCCA 374
Db 110648 GGGCTTTTGGCGGGGATGACGAGGAGCGGCCGACCTGTCTTCAGACGCGCCCA 110707

QY 375 CCTCTCTGATGGCAATTAATTAATTCGTATGCTGAATT 415
Db 110708 CCTCTCTGATGGCAATTAATTAATTCGTATGCTGAATT 110748

RESULT 4
LOCUS      I66494      7218 bp      DNA      PAT      28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION  I66494
VERSION     I66494.1 GI:2724471
KEYWORDS
SOURCE      unknown.
ORGANISM    unknown.
REFERENCE   1 (bases 1 to 7218)
AUTHORS     Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE       Recombinant foxpox virus
JOURNAL     Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source      1..7218
            /organism="unknown"
BASE COUNT  1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match      10.2%; Score 42.2; DB 5; Length 7218;
Best Local Similarity 2.8%; Pred. No. 0.033;
Matches 11; Conservative 214; Mismatches 162; Indels 0; Gaps 0;

QY 4 GGGACTTCAGTCTCCATCCATCCAGGAGCGAGTGGCCACTATGGGTCTGGGCTGCC 63
Db 1055 GGAGCTTGCATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1114

QY 64 CTTTCTCTCTCTTGACCTCTCTGGCAGTCTACATGGAACAGCGCGGGGTATGACTTT 123
Db 1115 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1174

QY 124 GCACTGAAGCTGAAGGAGCTTTTCTGACAAATCTCTCTATGATGTCAGCTTCGGA 183
Db 1175 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1234

QY 184 ATTGCTTGAAGAAAGCTCTGCTCTCTCTCCATCTCCCTTCAGGAGCAGCGTCAACCTCCA 243

```


TITLE

TITLE	HTGS Submission

Unpublished
2 (bases 1 to 65870)
Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M., Buettner, J., Butler, C., Card, P., deSailboat, C., Dunn, J., English, C., Ethridge, S., Garner, H.R., Gordon, M., Grant, O., Hamner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N., Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J., Sclutz, R.A., Stinson, S., Wagner, N., Waller, K. and Ward, T.
Direct Submission
Submitted (10-Oct-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
3 (bases 1 to 65870)
Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M., Buettner, J., Butler, C., Card, P., deSailboat, F., Dunn, J., English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G., Grant, O., Hamner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N., Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J., Sclutz, R.A., Stinson, S., Waller, K. and Ward, T.
Direct Submission
Submitted (01-Dec-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
On Dec 1, 1998 this sequence version replaced gi:3738098.
Location/Qualifiers
l . 65870
/organism="Mus musculus"
/db_xref="taxon:10090"
17024 a 15739 c 14910 g 18197 t

Query Match 8.7%; Score 36.2; DB 12; Length 65870;
Best Local Similarity 50.3%; Pred. No. 2.4;
Matches 89; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 176 TTCTGTGGAGTTCGTTGAAGAAGCTCGCCTCTCCATCTGCCCTCAGGACACCGCTC 235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30632 TCCTTGCGCTCTCTAGCTCTCTCTTGTCATATCCATCTCAACAATCGGGCCACCCTG 30573
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 ACCCTGCACCATCAAGATCTCAACACCATGTTGTCGCAACATCACAGGACATTGAAG 295
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30572 TCCTTCATCATGATCATGACGCTCATATCTCAGAGTATCCATCTCTATCTGTCCAATAAG 30513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 296 CTCTGTGCTCTTCTGCGCGGGCTTTGGCGGGGATGTCAGGAGGACGAGCCCCGAC 352
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30512 CAGETTTTCYTATAGCCGACTACTATATGCGGGGATTCCTGAGGGCGCTGGCCGAC 30456
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
AC006329/c
LOCUS
DEFINITION Homo sapiens clone NH0132A01, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION AC006329
VERSION AC006329.2 GI:4415968
KEYWORDS HWI: HGVS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 207984)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
2 (bases 1 to 207984)
AUTHORS Waterston,R.H.
TITLE Direct Submission
Submitted (09-JAN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
ON Mar 13, 1999 this sequence version replaced gi:4138779.
* NOTE: This is a 'working draft' sequence. It currently

```

* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

FEATURES	source
1	26566: contig of 2656 bp in length
1	2657
1	2658: gap of unknown length
1	2685
1	8135: contig of 3451 bp in length
1	8136
1	8135: gap of unknown length
1	8134
1	14419: contig of 6266 bp in length
1	14420
1	14437: gap of unknown length
1	14438
1	25405: contig of 10368 bp in length
1	25406
1	25423: gap of unknown length
1	25424
1	35744: contig of 10321 bp in length
1	35745
1	35762: gap of unknown length
1	35763
1	62631: contig of 26869 bp in length
1	62632
1	62649: gap of unknown length
1	62650
1	84733: contig of 22084 bp in length
1	84734
1	84751: gap of unknown length
1	84752
1	136772: contig of 52021 bp in length
1	136773
1	136790: gap of unknown length
1	136791
1	207984: contig of 71194 bp in length.
1	Location/Qualifiers
1	207984
1	/organism="Homo sapiens"
1	/db_xref="taxon:9606"
1	/clone="NH0132A01"
BASE COUNT	53798 a 51576 c 52041 g 50425 t 144 others
ORIGIN	

```

Query Match      8.5%; Score 35.4; DB 33; Length 207984;
Best Local Similarity 49.7%; Pred. No. 4.5;
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```

QY	100	TGCACAGCGGCCGGTATGACATTGCCACGTGAAGCTGAAGGAGTCTTTCTGACAAATTC	159
Db	5064	TGAGCAGAGCTGGCGGGATTTTGGCAGCCCGCTCTTAGGAGAAATGTTTGGGAAGACTC <td>5005</td>	5005
QY	160	CTCCTATGAGTCCAGCTCTCCTGGAAATGCTTGAAGAAGCTCTGGCTCTCTCTCCATCTCCC	219
Db	5004	CTTTTCCACAAAGCATCCCGGCCGCGAGCGAGCCCATCTGCTGGGCGCTTAATCTCGT	4945
QY	220	TTCCAGGACAGGCTCACCCCTCCACCATGCCAGATCTCAACACCATGTTGTGTCGAACAC	279
Db	4944	CTGAGAGCCCTGGAAATCCACCCACAGAACATCTTCTACATCCCTTGCTTCAAGC	4885

RESULT	9
ACOL13564	
TAXOCUS	DNA
DEFINITION	Homo sapiens chromosome 15 clone RP11-321f6 map 15, WORKING DRAFT SEQUENCE, 19 unordered pieces.
ACOL13564	
ACCESSION	AC013564.2 GI:6514010
VERSION	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT. human.
SOURCE	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 207092)
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Homo sapiens chromosome 15, clone RP11-321f6
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 207092) Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M, Balwin,J., Barna,N., Beckerly,R., Boquistaviv.L., Roukhaeifer,B

•

<http://genome.wustl.edu/gsc>
mailto:sapiens@watson.wustl.edu

NOTICE: This 'working draft' quality sequence may consist of several contigs from automated sequence assembly concatenated together. No attempt has been made to order or orient the contigs relative to one another correctly before concatenating. At each location in the sequence where contigs have been joined, several consecutive Ns may have been inserted.

The attached annotation was produced using a purely automated procedure.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

FEATURES

Source

1..148231

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="unknown"

/clone="NH0163G21"

/clone_lib="unknown"

1840..1961

/rpt_family="MIR"

2039..2122

/rpt_family="MIR"

2196..2601

/rpt_family="L1"

3563..3714

/rpt_family="MIR"

3947..4338

/rpt_family="L2"

5774..5951

/rpt_family="(TGAG)n"

8803..8851

/rpt_family="MIR"

9022..9213

/rpt_family="MIR"

9402..10001

/rpt_family="MER4-group"

10353..11049

/note="pseudogene similar to A28996 (PID:g91208)"

proline-rich protein M14 precursor - mouse"

11110..11202

/rpt_family="MaLR"

11951..12041

/rpt_family="MIR"

12532..12660

/rpt_family="MIR"

12697..12719

/rpt_family="(TTTA)n"

12720..13002

/rpt_family="Alu"

14466..14509

/rpt_family="T-rich"

14511..15076

/rpt_family="L1"

15804..15829

/rpt_family="(TTCA)n"

16282..16416

/rpt_family="Retroviral"

16750..16798

/rpt_family="(TA)n"

16798..16842

/rpt_family="(CA)n"

19964..19986

/rpt_family="AT-rich"

20606..20663

/rpt_family="Retroviral"

20665..20723

repeat_region /rpt_family="(TA)n"
21232..21253
repeat_region /rpt_family="AT-rich"
21415..21876
repeat_region /rpt_family="L2"
22093..22165
repeat_region /rpt_family="MIR"
23940..24094
repeat_region /rpt_family="(TG)n"
24163..24325
repeat_region /rpt_family="(TATATG)n"
24304..24451
repeat_region /rpt_family="(TG)n"
24462..24524
repeat_region /rpt_family="MER4-group"
24525..25055
repeat_region /rpt_family="(TG)n"
25328..25695
repeat_region /rpt_family="MaLR"
25931..26026
repeat_region /rpt_family="L2"
27904..27962
repeat_region /rpt_family="GA-rich"
28278..28730
repeat_region /rpt_family="L2"
28917..29288
repeat_region /rpt_family="MaLR"
30027..30128
repeat_region /rpt_family="MIR"
32576..32706
repeat_region /rpt_family="MIR"
33879..34233
repeat_region /rpt_family="MER21-group"
34234..34593
repeat_region /rpt_family="MaLR"
34594..35003
repeat_region /rpt_family="MER21-group"
35566..36099
repeat_region /rpt_family="MIR"
36805..36841
repeat_region /rpt_family="Alu"
37195..37254
exon 38143..38283
exon 38442..38696
repeat_region /rpt_family="G-rich"
39567..39779
misc_feature /note="pseudogene similar to P35085 (PID:g461699) CALCION
BINDING PROTEIN"
40070..40288
repeat_region /rpt_family="MER2_type"
42643..42960
44800..45117
misc_feature /note="pseudogene similar to PID:g1070603 collagen alpha
1(III) chain precursor - human"
45696..45752
repeat_region /rpt_family="L2"
46140..46559
repeat_region /rpt_family="L2"
46918..47149
repeat_region /rpt_family="MaLR"
47503..47571
repeat_region /rpt_family="MIR"
49216..49583
repeat_region /rpt_family="MaLR"
49866..50010
repeat_region /rpt_family="MIR"
51046..51183
repeat_region /rpt_family="L2"
51453..51806
repeat_region /rpt_family="MaLR"
51807..52111
repeat_region /rpt_family="Alu"
52112..52185

```

repeat_region /rpt_family="MELR"
52475..52955
repeat_region /rpt_family="L2"
53870..54109
repeat_region /rpt_family="MALR"
54110..54357
repeat_region /rpt_family="ALU"
54358..54418
repeat_region /rpt_family="MALR"
54419..54720
repeat_region /rpt_family="ALU"
54721..54803
repeat_region /rpt_family="MALR"
55004..55659
repeat_region /rpt_family="L1"
55769..55940
repeat_region /rpt_family="MIR"
55944..56399
repeat_region /rpt_family="MALR"
57034..57065
repeat_region /rpt_family="AT-rich"
57430..57728
repeat_region /rpt_family="ALU"

Query Match      8.4%  Score 34.8; DB 42; Length 148231;
Best Local Similarity 43.5%; Pred. No. 6.6;
Matches 108; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 141 AGCTTTTCGACAAATCCCTCATGATGTCAGTCCAGCTTCTCGAATTCCTTGAAGCTCT 200
Db 68150 AATATCTGTTTCAAAATCCCTGTTGACCAAGATATGAGGAGCAGTTCGAGATTCNN 68209
QY 201 GCCTCCTCCCTCCCTCCCTTCAGGAGCAGAGGTCACCTCCACCATGACGATCTCAAC 260
Db 68210 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 68269
QY 261 ACATGTTGTCACACATGACGATGAGCCTGTCCTCTTGGCCCGGGCTT 320
Db 68270 TGTAGTATGTCAGGTCGCCCTCTCTGCGGAGGCGCGTGTCTCCAGTGCCACGGTT 68329
QY 321 TTGGCGCGGGATGACAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
Db 68330 CTTGCTGCGGATGACTCTGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68389
QY 381 TGAGTGGC 388
Db 68390 TGTGAGC 68397

RESULT 11
AC012184/C
LOCUS Homo sapiens chromosome 16 clone RPI-11_529K1, 21-OCT-1999
DEFINITION PROGRESS ***, 40 unordered pieces.
ACCESSION AC012184
VERSION AC012184.1 GI:6091684
KEYWORDS HTG; HIGS-PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 65673)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 65673)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 760: contig of 760 bp in length
 gap of unknown length
 761 1560: contig of 800 bp in length
 gap of unknown length
 1561 2303: contig of 743 bp in length
 gap of unknown length
 2304 3255: contig of 952 bp in length
 gap of unknown length
 3256 4246: contig of 991 bp in length
 gap of unknown length
 4247 5603: contig of 1357 bp in length
 gap of unknown length
 5604 6762: contig of 1159 bp in length
 gap of unknown length
 6763 7004: contig of 242 bp in length
 gap of unknown length
 7005 8046: contig of 1042 bp in length
 gap of unknown length
 8047 9309: contig of 1263 bp in length
 gap of unknown length
 9310 10988: contig of 1679 bp in length
 gap of unknown length
 10989 12037: contig of 1049 bp in length
 gap of unknown length
 12038 13400: contig of 1363 bp in length
 gap of unknown length
 13401 14099: contig of 699 bp in length
 gap of unknown length
 14100 15399: contig of 1300 bp in length
 gap of unknown length
 15400 16772: contig of 1373 bp in length
 gap of unknown length
 16773 17868: contig of 1096 bp in length
 gap of unknown length
 17869 18643: contig of 775 bp in length
 gap of unknown length
 18644 19788: contig of 1145 bp in length
 gap of unknown length
 19789 21845: contig of 2057 bp in length
 gap of unknown length
 21846 23223: contig of 1378 bp in length
 gap of unknown length
 23224 25010: contig of 1787 bp in length
 gap of unknown length
 25011 25466: contig of 456 bp in length
 gap of unknown length
 25467 27098: contig of 1632 bp in length
 gap of unknown length
 27099 28274: contig of 1176 bp in length
 gap of unknown length
 28275 29748: contig of 1474 bp in length
 gap of unknown length
 29749 31363: contig of 1615 bp in length
 gap of unknown length
 31364 33150: contig of 1787 bp in length
 gap of unknown length
 33151 36067: contig of 2917 bp in length
 gap of unknown length
 36068 37581: contig of 1514 bp in length
 gap of unknown length
 37582 39288: contig of 1707 bp in length
 gap of unknown length
 39289 42139: contig of 2851 bp in length
 gap of unknown length
 42140 44809: contig of 2670 bp in length
 gap of unknown length
 44810 47022: contig of 2213 bp in length

Kubota, Osaka University Medical School, Department of Obstetrics and Gynecology; 2-2 Yamadaoka, Suita, Osaka 565, Japan (E-mail:kubota@gyne.med.osaka-u.ac.jp, tel:81-6-879-3351, Fax:81-6-879-3359)

2 (bases 1 to 3626)

Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Saiji, F. and Murata, Y.

Structure and Expression of the Mouse Oxytocin Receptor Gene Unpublished (1996)

3 (sites)

Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Saiji, F. and Murata, Y.

Structure and expression of the mouse oxytocin receptor gene Mol. Cell. Endocrinol. 124 (1-2), 25-32 (1996)

97179034

Location/Qualifiers

1. 3626

/organism="Mus sp."

/strain="129SV"

/db_xref="taxon:10095"

/cell_type="ES cell"

1. 1530

1531..1798

/number=1

1799..2240

/number=1

2241..2393

/number=2

2394..2485

/number=2

2486..3547

/number=3

2629..33546

/codon_start=1

/product="oxytocin receptor"

/protein_id="BAAL8996.1"

/db_xref="GI:1902966"

translating="MSGTPAANWSTELDLSGVPPGAGNLTAGPRNEALARVEVA

VICLLIFALSGNACVLLALPTTRKHSRLFFMKHLSADLVAVVFOVLPOLLMDIT

FRFGPDLLCRNLYKLYOVGMFASYLLLLKSLDCLALCOPLSRRLDRILAVLA

WGLCLSVSPQVHIFSLREAVDGVDFCNVFIQPNCPYATVITLAVIVPVLAA

CYGLSFKIWNQRLRTAAARAGSDAAGAGAGAAALRVSSVLLSKAKATVMT

FIIVLAFIVCWTFPTFFVQWMSVMDVNAPE"

3548..3626

/number=3

782 a 921 c 1131 g 792 t

BASE COUNT

ORIGIN

Query Match 8.2%; Score 34.2; DB 12; Length 3626;

Best Local Similarity 60.0%; Pred. No. 7.8;

Matches 57; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 152 ACAATTCCTCTGATGAGTCCAGCTTCCTGGAAATTCCTTGRAAGCTCTCCCTCTCTC 211

Db 2085 AAAAGCTTACTACGGGTACATCTCTCTGGGTCCGCCAAAGCCATCTCTCTC 2026

Oy 212 CATCTCCCTTCAGGACCGCTGACCTCCACCA 246

Db 2025 CTCCTGCTTCAAGACTCAGCCGCCAGCTCCCA 1991

RESULT 14

HSJ858B16

LOCUS

DEFINITION

Human DNA sequence from clone 858B16 on chromosome 22. Contains the KIAA0542 gene and a gene for a novel protein similar to hamster PSSC (phosphatidylserine decarboxylase proenzyme, EC 4.1.1.65).

Contains ESTs, GSSs and a putative CpG island, complete sequence.

AL096768

AL096768.7 GI:5596770

ACCESSION

VERSION

HTG: CpG Island; KIAA0542; Phosphatidylserine Decarboxylase

KEYWORDS

Proenzyme; PSSC.

human.

SOURCE

ORGANISM

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 36676)

Barlow, K.

Direct Submission

Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jul 27, 1999 this sequence version replaced gi:5579004.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source database: Em, EMBL; Swi, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr22

858B16 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://nacpac.med.buffalo.edu/ VECTOR: pCYPAC2

IMPORTANT: this sequence is not the entire insert of clone 858B16. It may be shorter because we only sequence overlapping actions once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 694E4 (AL031255) is at 36572 in this sequence.

Location/Qualifiers

1. 36676

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22"

/clone_lib="RPCI-5"

/clone="RP5-858B16"

/join("<273..454,1413..1488,1978..2073,2538..2614,3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,11023..11199,11271..11524,12717..12948,14414..14546,14669..14747,15784..15895,15983..16217")

/gene="dJ858B16.1"

/note="match: ESTs: Em:R53362 Em:278408 Em:AA521233 Em:R53564 Em:AI806913 Em:RA158770 Em:AA836285 Em:AA974403 Em:AA594535 Em:AI858196 Em:AI425026 Em:R51102 Em:AA582394 Em:AA899254 Em:AA553345 Em:AI809747 Em:RA84853 Em:AA971234 Em:R54651 Em:AI141022 Em:AA159572 Em:AI361548 Em:AA916775 Em:AI191287 Em:AV031353 Em:AA797944 Em:N64034 Em:AV075750 Em:AA095355 Em:AA865102 Em:AI232370 Em:AA595561 Em:AA088219 Em:W25871 Em:N56896 Em:W90666 Em:AV030271 Em:AV116402 Em:AV122327 Em:AI024249 Em:AA673371 Em:AI190893 Em:AI019759 Em:AI770152 Em:AA419582 Em:N30173 Em:AI342072 Em:R54847 Em:AI126154 Em:AI181158 Em:AI024124 Em:AA54928 Em:AA215582 Em:AA398600 Em:AI623440 Em:AI024124 Em:AA637284 Em:AI824779 Em:AI094871 Em:AI192847 Em:N63934 Em:AV083921

/evidence="not experimental"

/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"

join("<273..454,1413..1488,1978..2073,2538..2614,3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,11023..11199,11271..12948,14414..14546,14669..14747,15784..15895,15983..16218")

mrna


```

/feature="dJ858B16.1"
/feature="match: cDNAs: Em:AB011114"
/feature="not_experimental"
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
273..16218
/feature="dJ858B16.1"
join(273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12717..12948,14414..14546,
14659..14747,15784..15895,15983..16087)
/feature="dJ858B16.1"
/feature="different from published protein but supported by
GENES and GENSCAN"
/codon_start=1
/feature="not_experimental"
/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"
/protein_id="CAB56392.1"
/db_xref="GI:5921489"
/translators="OAILHAEROLLYRSWTFWQHQAARHQEQEMQVACAHHRGL
KKAFLRESAQLRTERTGRVAAEFHMAQLLWAWSQRECLALGAEKRLRAD
LHQHVSILHRAQWVYQGRVRSILREVAARESQHQLLRGALRWKENTMARVDE
AKKTFQASTHYRTICSKVLVQWREAVSVQMYRQEDCAIWEAQKVLDRGCLRTWFO
RWDCSRSQAQQLQARQORATVRALNFAPSLQAKVWATWLFVLRERKEALQ
RLSTCFRWRQOLARQORATVRALNFAPSLQAKVWATWLFVLRERKEALQ
WLAGYQGLQLEGATRLIRFAASMKASRQOLQAOQVQAHSIHRVRCATLWKQ
WLAGYQGLQLEGATRLIRFAASMKASRQOLQAOQVQAHSIHRVRCATLWKQ
PEPALGPKOPKOEGLGMAQAPASLRTFLAEAPTALVPHSPGLGALGLAA
PTASTGPELLLPLSFPCCGAARPARVSRQATPRDKPPVPSSLASVPDPHLLPGD
PSATRAGCLSTAGSLDEALEBETIOQLLHYQTTKQNLWSCRQASRLRWLENGE
EPGDEDEQQVQKELEQVEMQIQLAEELAQORQPTGACVARIQALQALC"
join(273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11528)
/feature="dJ858B16.1"
/feature="original published form; could be artefactual as it
is result of a large intron read-through; match: proteins:
Tr:060289"
/codon_start=1
/feature="not_experimental"
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
/protein_id="CAB56393.1"
/db_xref="GI:5921490"
/translators="OAILHAEROLLYRSWTFWQHQAARHQEQEMQVACAHHRGL
KKAFLRESAQLRTERTGRVAAEFHMAQLLWAWSQRECLALGAEKRLRAD
LHQHVSILHRAQWVYQGRVRSILREVAARESQHQLLRGALRWKENTMARVDE
AKKTFQASTHYRTICSKVLVQWREAVSVQMYRQEDCAIWEAQKVLDRGCLRTWFO
RWDCSRSQAQQLQARQORATVRALNFAPSLQAKVWATWLFVLRERKEALQ
RLSTCFRWRQOLARQORATVRALNFAPSLQAKVWATWLFVLRERKEALQ
WLAGYQGLQLEGATRLIRFAASMKASRQOLQAOQVQAHSIHRVRCATLWKQ
WLAGYQGLQLEGATRLIRFAASMKASRQOLQAOQVQAHSIHRVRCATLWKQ
VLRGGKPPQPLAIAIPSRKVTFFEGPLNIRIAGADGTLTKRPQASRLPLGALGLAA
EEPALEL"
620..913
/feature="AluX repeat: matches 1..294 of consensus"
1444..1687
/feature="dJ858B16.1"
/feature="match: GSS: Em:AQ545234"
3250..3336
/feature="AluJo repeat: matches 2..302 of consensus"
3634..3836
/feature="L2 repeat: matches 2497..2701 of consensus"
4356..4646
/feature="AluX repeat: matches 21..312 of consensus"
4650..4754
/feature="AluSg/x repeat: matches 200..307 of consensus"
4900..5281
/feature="AluX repeat: matches 1..292 of consensus"
5806..6231
/feature="MLTIC repeat: matches 6..460 of consensus"
6529..6668
/feature="FRAM repeat: matches 1..140 of consensus"
6671..6966
/feature="AluY repeat: matches 1..295 of consensus"
6985..7081

```

```

/feature="LTP40a repeat: matches 1..100 of consensus"
7929..8097
/feature="MIR repeat: matches 82..255 of consensus"
9766..10068
/feature="AluY repeat: matches 1..301 of consensus"
10069..10096
/feature="14 copies 2 mer ta 100% conserved"
10104..10399
/feature="AluJb repeat: matches 1..293 of consensus"
12099..12259
/feature="MIR repeat: matches 46..208 of consensus"
13063..13779
/feature="L2 repeat: matches 1570..2750 of consensus"
15108..15404
/feature="AluSg repeat: matches 1..295 of consensus"
complement(16160)
/feature="dJ858B16.2"
complement(join(16160..17505,18222..18382,18665..18811,
28473..28494))
/feature="dJ858B16.2"
/feature="match: cDNAs: Em:M62722 Em:AL050371 Em:AF086277
Em:AF035304 Em:U92163 Em:AB011114; match: ESTs:
Em:AA324939 Em:AA582112 Em:AA166445 Em:AA166440
Em:AA360704 Em:AI019759 Em:AL039237 Em:AA118158
Em:AV083921 Em:AI232370 Em:AU035550 Em:AV12327
Em:AA171991 Em:AA326672 Em:AU051210 Em:AA288268
Em:AA144920 Em:AA673371 Em:R00603 Em:AA297077 Em:AA237290
Em:AA615772 Em:AA297754 Em:AA297076 Em:AA643609
Em:AI094871 Em:AA298523 Em:AA149582 Em:AA836285
Em:AV031353 Em:AA297159 Em:AI191287 Em:AA916775 Em:R51102
Em:N30173 Em:AA521233 Em:AA044074 Em:AI361548 Em:H69648
Em:AA971234 Em:W00857 Em:AI623440 Em:AA553345 Em:H93993
Em:AA046738 Em:AA637284 Em:AA151798 Em:W69452 Em:AA297086
Em:W69536 Em:T10522 Em:AT024124 Em:H71693 Em:AA302951
Em:N98260 Em:N63934 Em:AA215768 Em:R94341 Em:AA177342
Em:R48874 Em:AI342072 Em:W90666 Em:AA255719 Em:AI042429
Em:AA829254 Em:W69460 Em:AA595561 Em:W69544 Em:T77685
Em:AA779744 Em:N64034 Em:AV075750 Em:H38339 Em:AV095355
Em:AI16402 Em:AI425026 Em:AA974403 Em:AA908164
Em:AA088219 Em:AA298551 Em:N56896 Em:AV030271 Em:AI428571
Em:AI806913 Em:AI014094 Em:H53175 Em:AI809747 Em:H38743
Em:H62778 Em:W03262 Em:R52264 Em:AA044185 Em:R82229
Em:AA369042 Em:AA393183 Em:AI126154 Em:R54847 Em:AA158770
Em:AA825239 Em:R53399 Em:AI770152 Em:AA594535 Em:AA582394

```

Query Match 8.2%; Score 34; DB 11; Length 36676;
Best Local Similarity 54.9%; Pred. No. 10;

Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```

Qy 15 GTCTCTCCATCCAGGAGCGAGTGGCCATGTTGGGTCTGGGTCTGGCTTGTCTCTCC 74
Db 14428 GGCTACTCTAGGAGTAAGCCCGGCTCCCTCATCCCTGGCCAGTGTCTCTGACCCCA 14487

Qy 75 TCTTGACCTCTTGGACGCTCACATGGRACAGCGCGCGGTATGCTTTCGCACTGAC 134
Db 14488 TCTACTCTCTCTGGGACTTCTAGCCACAGGCTGGGCTTTCGCACTTTCGCACTG 14547

Qy 135 TG 136
Db 14548 TG 14549

```

RESULT 15

```

AC016298
LOCUS AC016298 62181 bp DNA HTG 24-NOV-1999
DEFINITION Homo sapiens Clone RP11-655M17, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016298
VERSION AC016298.1 GI:6467036
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 62181)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-655M17
Unpublished
2 (bases 1 to 62181)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castelle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galgani,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,K., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submission@genome.wi.mit.edu
----- Project Information
Center project name: I4315
Center clone name: 655_M17
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 775: contig of 775 bp in length
* gap of unknown length
* 776 1583: contig of 808 bp in length
* gap of unknown length
* 1584 2371: contig of 788 bp in length
* gap of unknown length
* 2372 3185: contig of 814 bp in length
* gap of unknown length
* 3186 3980: contig of 795 bp in length
* gap of unknown length
* 3981 4778: contig of 798 bp in length
* gap of unknown length
* 4779 5584: contig of 806 bp in length
* gap of unknown length
* 5585 6392: contig of 808 bp in length
* gap of unknown length
* 6393 7189: contig of 797 bp in length
* gap of unknown length
* 7190 7994: contig of 805 bp in length
* gap of unknown length
* 7995 8802: contig of 808 bp in length
* gap of unknown length
* 8803 9603: contig of 801 bp in length
* gap of unknown length
* 9604 10400: contig of 797 bp in length
* gap of unknown length
* 10401 11206: contig of 806 bp in length
* gap of unknown length
*
* 11207 12024: contig of 818 bp in length
* gap of unknown length
* 12025 12840: contig of 816 bp in length
* gap of unknown length
* 12841 13637: contig of 797 bp in length
* gap of unknown length
* 13638 14440: contig of 803 bp in length
* gap of unknown length
* 14441 15234: contig of 794 bp in length
* gap of unknown length
* 15235 16031: contig of 797 bp in length
* gap of unknown length
* 16032 16836: contig of 805 bp in length
* gap of unknown length
* 16837 17615: contig of 779 bp in length
* gap of unknown length
* 17616 18421: contig of 806 bp in length
* gap of unknown length
* 18422 19233: contig of 812 bp in length
* gap of unknown length
* 19234 20042: contig of 809 bp in length
* gap of unknown length
* 20043 20856: contig of 814 bp in length
* gap of unknown length
* 20857 21657: contig of 801 bp in length
* gap of unknown length
* 21658 22462: contig of 805 bp in length
* gap of unknown length
* 22463 23262: contig of 800 bp in length
* gap of unknown length
* 23263 24065: contig of 803 bp in length
* gap of unknown length
* 24066 24866: contig of 801 bp in length
* gap of unknown length
* 24867 25673: contig of 807 bp in length
* gap of unknown length
* 25674 26475: contig of 802 bp in length
* gap of unknown length
* 26476 27270: contig of 795 bp in length
* gap of unknown length
* 27271 28075: contig of 805 bp in length
* gap of unknown length
* 28076 28817: contig of 742 bp in length
* gap of unknown length
* 28818 29613: contig of 796 bp in length
* gap of unknown length
* 29614 30416: contig of 803 bp in length
* gap of unknown length
* 30417 31207: contig of 791 bp in length
* gap of unknown length
* 31208 32021: contig of 814 bp in length
* gap of unknown length
* 32022 32825: contig of 804 bp in length
* gap of unknown length
* 32826 33591: contig of 786 bp in length
* gap of unknown length
* 33592 34403: contig of 812 bp in length
* gap of unknown length
* 34404 35213: contig of 810 bp in length
* gap of unknown length
* 35214 36022: contig of 809 bp in length
* gap of unknown length
* 36023 36829: contig of 807 bp in length
* gap of unknown length
* 36830 37623: contig of 794 bp in length
* gap of unknown length
* 37624 38411: contig of 788 bp in length
* gap of unknown length
* 38412 39213: contig of 802 bp in length
* gap of unknown length
* 39214 40010: contig of 797 bp in length
* gap of unknown length

```

TITLE

JOURNAL

COMMENT

* 40011 40741: contig of 731 bp in length
* gap of unknown length
* 40742 41570: contig of 829 bp in length
* gap of unknown length
* 41571 42298: contig of 728 bp in length
* gap of unknown length
* 42299 42980: contig of 682 bp in length
* gap of unknown length
* 42981 43786: contig of 806 bp in length
* gap of unknown length
* 43787 44584: contig of 798 bp in length
* gap of unknown length
* 44585 45392: contig of 808 bp in length
* gap of unknown length
* 45393 46206: contig of 814 bp in length
* gap of unknown length
* 46207 46991: contig of 785 bp in length
* gap of unknown length
* 46992 47813: contig of 822 bp in length
* gap of unknown length
* 47814 48623: contig of 810 bp in length
* gap of unknown length
* 48624 49437: contig of 814 bp in length
* gap of unknown length
* 49438 50183: contig of 746 bp in length
* gap of unknown length
* 50184 50985: contig of 802 bp in length
* gap of unknown length
* 50986 51800: contig of 815 bp in length
* gap of unknown length
* 51801 52595: contig of 795 bp in length
* gap of unknown length
* 52596 53395: contig of 800 bp in length
* gap of unknown length
* 53396 54205: contig of 810 bp in length
* gap of unknown length
* 54206 54996: contig of 791 bp in length
* gap of unknown length
* 54997 55783: contig of 787 bp in length
* gap of unknown length
* 55784 56587: contig of 804 bp in length
* gap of unknown length
* 56588 57389: contig of 802 bp in length
* gap of unknown length
* 57390 58203: contig of 814 bp in length
* gap of unknown length

Query Match 8.1% Score 33.8; DB 44; Length 62181;
Best Local Similarity 53.4%; Pred. No. 12;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCTGCC 65
Db 11905 GATTTTAATTCCTCTACCCCTGAAGTGGAGTGGTCTCTCTTTGGTGGCTCCCC 11964
QY 66 TTGTCCTCTCTTGACCTCTCTGGCAGCTCACATGGACAGGCGCGGTATGACTTGC 125
Db 11965 TCTTCTTCTTTGGCCCTGGGTGGCAGTCCCTGGAAPANGGTCTCAATCCCTTGC 12024
QY 126 AACTGAGCTGAA 138
Db 12025 GCCTCGGGTTGCA 12037

Search completed: March 18, 2000, 20:24:08
Job time: 1518 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2000, 09:31:54 ; Search time 76.27 Seconds
(without alignments)
1361.344 Million cell updates/sec

Title: US-09-092-296-7_COPY_5_419

Perfect score: 415

Sequence: 1 ACCGGAGCTTCAGTGTCTCC.....AAAATGGTATGCTGAATT 415

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	411	99.0	439	1	V84366	Human stomach carc
2	34.4	8.3	2923	1	X04326	Human secreted pro
3	32.2	7.8	110000	1	X20248_03	Continuation (4 of
4	32	7.7	1128	1	V28846	Mouse coxsackievir
5	31	7.5	4051	1	V62672	Tumour necrosis fa
6	31	7.5	3964	1	V62673	Cell-cycle regulat
7	30.4	7.3	580	1	V20964	Mouse coxsackievir
8	30.4	7.3	1301	1	V50430	Human TR6 cDNA, DN
9	30.2	7.3	3881	1	V63094	Human interleukin-
10	30	7.2	7824	1	V74052	Human gene signatu
11	30	7.2	287	1	V73981	EST clone AR34. Ne
12	30	7.2	14590	1	X22303	HAP-1 radiation pr
13	29.8	7.2	3310	1	T18696	Polynucleotide seq
14	29.8	7.2	291	1	V86365	Coding sequence fo
15	29.8	7.2	20757	1	X20599	Simian Transformin
16	29.4	7.1	6175	1	T34291	Human transforming
17	29.2	7.0	1560	1	Q03508	Simian transforming
18	29.2	7.0	1561	1	N00000	Coding sequence of
19	29.2	7.0	1571	1	Q03269	Human transforming
20	29.2	7.0	1561	1	Q03268	Human transforming
21	29.2	7.0	1560	1	X81084	TATA-binding prote
22	29.2	7.0	5862	1	Q70729	Human TATA-binding
23	29.2	7.0	5862	1	T42215	Human TATA-binding
24	29.2	7.0	5862	1	T79596	Monkey transformin
25	29.2	7.0	1303	1	Q09317	Sequence encoding
26	29	7.0	2259	1	Q05106	Human apolipoprote
27	29	7.0	1110	1	N60409	Sequence encoding
28	29	7.0	1110	1	N50450	Human apolipoprote
29	29	7.0	597	1	Q69099	Sequence encoding
30	29	7.0	936	1	Q69101	Human ApoE4 cDNA.
31	29	7.0	597	1	T18068	Human ApoE4Lx2 cDN
32	29	7.0	936	1	T18070	Human apolipoprote
33	29	7.0	1157	1	T06957	Human apolipoprote
34	29	7.0	3060	1	T62781	Urea transporter p
35	29	7.0	3580	1	H81742	Human PT04 encodi
36	28.8	6.9	1785	1	T66986	Human brain-specif
37	28.8	6.9	488	1	V11943	HIV-1 sub-type B g
38	28.8	6.9	844	1	V34155	Human secreted pro
39	28.4	6.8	6306	1	Q54841	Sequence encoding

c 40 28.4 6.8 2262 1 V06554 Alcaligenes (Deley
c 41 28 6.7 5018 1 N70646 Clone 2386p102 enc
c 42 28 6.7 5018 1 R50150 Sequence of the cd
43 28 6.7 5350 1 T89254 cDNA for rat CASPR
44 28 6.7 795 1 V34224 Human secreted pro
45 28 6.7 325 1 V90470 EST clone DT544. N

ALIGNMENTS

RESULT 1
V84366
ID V84366 standard; cDNA to mRNA; 439 BP.
AC V84366;
DT 30-MAR-1999 (first entry)
DE Human stomach carcinoma cDNA clone HP10408.
KW Transmembrane protein; HP10408; human; stomach cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 75..311
FT /*tag= a
FT /note= "cDNA comprising the coding region (minus
the stop codon) is claimed (Claim 3)"
FN W09855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI: 99-045730/04.
DR P-PSDB: W88498.
PT New human proteins containing transmembrane domains and their
encoding sequences - useful in the preparation of antibodies and
large-scale protein production, gene diagnosis, and gene therapy
Claim 4: Page 135; 178pp; English.
CC This is the nucleotide sequence of cDNA clone HP10408, which
includes a coding region (also claimed) for a novel human
transmembrane protein (see W88498). The clone was isolated from a
stomach cancer cDNA library using a signal sequence detection
method, and by protein synthesis by in vitro translation. The
encoded protein has a putative signal sequence and a putative
internal transmembrane domain. The invention provides nucleotide
sequences (see W84359-76) coding for 18 transmembrane proteins
(see W88491-508), vectors containing such polynucleotides, and
eukaryotic cells containing the vectors. The proteins can be
used as antigens or as compositions in the preparation of
antibodies against the proteins. The polynucleotides can be used
as probes for gene diagnosis, and as gene sources for gene therapy
and large-scale production of proteins encoded by the cDNA. The
host cells are used for the detection of ligands corresponding to
the expressed proteins, and the screening of low mol.wt. medicines.
SQ Sequence 439 BP; 89 A; 137 C; 109 G; 104 T;

Query Match 99.08; Score 411; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.3e-116;
Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCGGAGCTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTTGGGCT 60
Db ACCGGAGCTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTTGGGCT 88
QY 61 GCCCCTGTCTCTCTCTGTACCTCTTGGCAGTCACATGGAGCCGCGGTATGAC 120
Db GCCCCTGTCTCTCTCTGTACCTCTTGGCAGTCACATGGAGCCGCGGTATGAC 148
QY 121 TTGCAACTGAAGCTGAAGGAGTCTTTTTCACAAATTCCTCTATGAGTCAGGCTTCT 180
Db TTGCAACTGAAGCTGAAGGAGTCTTTTTCACAAATTCCTCTATGAGTCAGGCTTCT 208
QY 181 GGAATGCTTGAAGAAGCTCTGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240

PR 16-APR-1997; US-843652.
 PI (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 DR Holtzman D;
 DR WPI; 98-594562/50.
 DR P-PSDB; W79261.
 PT Isolated tumour necrosis factor related proteins - used to develop
 PT products for the diagnosis and treatment of apoptosis-related
 PT disorders, e.g. cancers, autoimmune disorders or neurodegenerative
 PT disorders.
 PS Claim 7; Fig 2; 88pp; English.
 CC This cDNA sequence includes a coding region for a new member of the
 CC human tumour necrosis factor receptor superfamily that is designated
 CC Tango-63e (see W79261). 2 Different forms of Tango-63, i.e.
 CC Tango-63e and Tango-63d (see W62672), were identified in a human
 CC prostate cDNA library through EST sequencing and screening. The
 CC encoded polypeptides are identical with the exception of the
 CC deletion of amino acids 183-211 of Tango-63d (see W79260) in
 CC Tango-63e. The Tango-63 gene was mapped to chromosome 8 between
 CC markers WI-6088 and WI-6563. A plasmid encoding Tango-63e is
 CC deposited as ATCC 98367. The invention encompasses nucleic acid
 CC molecules encoding Tango-63d and -63e, vectors containing these
 CC nucleic acid molecules, cells harboring recombinant DNA encoding
 CC Tango-63d and/or -63e, fusion proteins that include Tango-63d
 CC and/or -63e, transgenic animals that express Tango-63d and/or -63e,
 CC and recombinant knockout animals that fail to express Tango-63d
 CC and/or -63e. Methods are provided for the diagnosis and treatment
 CC of disorders associated with either an abnormally high or an
 CC abnormally low rate of apoptotic cell death. Inhibitors can be
 CC used for treating e.g. cancers, autoimmune disorders (e.g. systemic
 CC lupus erythematosus and immune-mediated glomerulonephritis), and
 CC viral infections (e.g. herpesviruses, poxviruses, and adenoviruses).
 CC Agonists can be used for treating e.g. neurodegenerative diseases,
 CC e.g. Alzheimer's disease, Parkinson's disease, amyotrophic lateral
 CC sclerosis (ALS), Huntington's disease, retinitis pigmentosa, spinal
 CC muscular atrophy, various forms of cerebellar degeneration, anaemia,
 CC myelodysplastic syndrome, ischemic injury, myocardial infarction,
 CC cerebral ischemia or toxin-induced injury. In addition, T cell
 CC mediated diseases, including AIDS, autoimmune diseases such as
 CC rheumatoid arthritis, and type I diabetes, septic shock, cerebral
 CC malaria, graft rejection, cytotoxicity, cachexia, and inflammation
 CC can be treated by altering the expression or activity of the
 CC polypeptides. The products can also be used for detection,
 CC diagnosis and screening assays. 939 C; 995 G; 1042 T;
 SQ Sequence 3964 BP; 981 A; 939 C; 995 G; 1042 T;

Query Match 7.5%; Score 31; DB 1; Length 3964;
 Best Local Similarity 62.0%; Pred. No. 5.8;
 Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 233 GTCACCTCCACATGCTCAACACCATGTTGCTGCAACACATGACAGCCATG 292
 Db 3684 GTTACATAGGCTTTGCTATATCCCGCCAAATAGCATGTGACACAGGACGCATAG 3743

QY 293 AAGCCTGTGCTCTCTGG 311
 Db 3744 TATAGTGTCACTCGTGG 3762

RESULT 7
 ID T02964/c
 AC T02964;
 DT 01-MAR-1996 (first entry)
 DE Cell-cycle regulatory protein p15 cDNA.
 KW Cell-cycle regulatory protein p15; cyclin-dependent kinase inhibitor;
 KW CCR; gene therapy; transgenic animal; cancer; cell proliferation;
 KW ss; ds.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT cds 91..483
 FT /*tag= a
 PN WO9528483-AL.

PD 26-OCT-1995.
 PF 14-APR-1995; US-04636.
 PR 14-APR-1994; US-227371.
 PR 25-MAY-1994; US-248812.
 PR 14-SEP-1994; US-306511.
 PR 29-NOV-1994; US-346147.
 PA (COLD-) COLD SPRING HARBOR LAB.
 PI Beach DH, Demetriuk DJ, Hannon GJ, Serrano M;
 PI WPI; 95-373798/48.
 DR P-PSDB; R85118.
 DR New cell cycle regulating proteins bind to cyclin dependent kinase -
 PT and related nucleic acids, antibodies etc., used in diagnosis and
 PT therapy of abnormal cell proliferation, degeneration etc.
 PS Claim 43; Page 81-82; 109pp; English.
 CC cDNA (T02964) coding for the mouse cell-cycle regulatory (CCR)
 CC protein p15 (R85118) was isolated from an embryonal carcinoma
 CC library using a probe based on a mouse CCR p13.5 cDNA (T02965).
 CC The isolated cDNA can be used to detect mutations in CCR
 CC genes that lead to cell proliferation; to breed transgenic
 CC animals to study cellular disorders involving CCR allele
 CC mutation/misexpression; and to correct CCR-deficient cells
 CC (gene therapy).
 SQ Sequence 580 BP; 109 A; 187 C; 197 G; 87 T;

Query Match 7.3%; Score 30.4; DB 1; Length 580;
 Best Local Similarity 57.3%; Pred. No. 4.1;
 Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 35 GCGATGGCCACTATGGGTCGTGGCTGCTCCCTTGCTCCCTTGACCTCTCTGSCAGC 94
 Db 128 GCGTGGCCAGGCCGCGTCACTGCTCCGCCACACATGCTTGTCCTGCTGTGGC 69

QY 95 TCACATGAACAGGCCGCGGTATGACTTTGCAACTG 130
 Db 68 AGAAATGCTCTTCGCCGCGGTGAGATTGCTACAG 33

RESULT 8
 ID V50430
 AC V50430;
 DT 07-DEC-1998 (first entry)
 DE Mouse coxsackievirus and Ad2 and Ad5 receptor (MCAR) cDNA.
 KW MCAR; coxsackievirus receptor; CVB; adenovirus; Ad2 receptor;
 KW Ad5 receptor; mouse; infection; vaccine; gene therapy; ds.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT cds 157..1215
 FT /*tag= a
 PN WO9833819-AL.
 PD 06-AUG-1998.
 PF 30-JAN-1998; U01724.
 PR 30-JAN-1997; US-036986.
 PA (UNYNY) UNIV NEW YORK STATE.
 PI Philipson L, Tomko RP;
 DR WPI; 98-437397/37.
 DR P-PSDB; W69698.
 DT DNA encoding human receptor for adenovirus C and coxsackievirus B -
 PT for preventing and treating viral infection and rendering cells
 PT susceptible to transformation by adenoviral vectors in gene therapy
 PS Claim 2; Page 68-70; 88pp; English.
 CC This cDNA molecule codes for mouse MCAR protein (see W69698) that
 CC serves as a cellular receptor for adenoviruses of the serotypes 2
 CC and 5 (subgroup C) and for the group B coxsackieviruses (CVB). It
 CC was obtained by screening a lambda phage expression cDNA library
 CC with antiserum containing antibodies specific for mouse CVB
 CC binding protein p46, and was used to identify human HCAR cDNA
 CC (see V50429). The invention also provides host cells transformed
 CC with DNA molecules encoding HCAR or MCAR and methods of producing
 CC the recombinant proteins or their derivatives. These proteins,
 CC their extracellular domains, as well as oligopeptides (see
 CC W69699-708) which bind virus, are also provided. Isolated HCAR or


```
CC MCAR proteins or their fragments or variants are used to prevent
CC or treat virus infections. Expressing the DNA in cells which lack
CC these viral receptors renders the cells susceptible to
CC transformation by adenoviral vectors carrying genes for gene
CC therapy.
SQ Sequence 1301 BP; 355 A; 327 C; 322 G; 297 T;

Query Match 7.3%; Score 30.4; DB 1; Length 1301;
Best Local Similarity 61.2%; Pred. No. 5.7;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 29 AGGAGCGCAGTGGCCACTATGGGCTGTGGCTGGCCCTTGTCTCTTGACCCCTCTT 88
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 868 ACGATCGGGGCGCGTCAATAGGAGCGCTGTGGCCCTTGTCTCATCGGGCCATCCTC 927

QY 89 GGCAGCTCACATGGACAGG 108
    ||||| ||||| ||||| |||||
Db 928 TTCGTGTGTACAGGAACG 947

RESULT 9
V63094
ID V63094 standard; cDNA; 3881 BP.
AC V63094;1999 (first entry)
DE Human TR6 cDNA.
KW TR6; tumour necrosis factor related receptor; human; treatment; stroke;
KW inflammation; arthritis; septicaemia; autoimmune disease; restenosis;
KW transplant rejection; infection; ischaemia; brain injury; bone disease;
KW acute respiratory disease syndrome; acquired autoimmune disease syndrome;
KW AIDS; cancer; atherosclerosis; Alzheimers disease; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 94..1329
    /*tag= a
    /product= "TR6"
EP-870827-A2.
PD 14-OCT-1998.
PF 23-DEC-1997; 310562.
PR 22-AUG-1997; US-916625.
PR 14-MAR-1997; US-041230.
PR 09-MAY-1997; US-953684.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Deen KC, Young PR;
DR WPI; 98-523156/45.
P-PSDB; W76827.
DR DNA encoding tumour necrosis factor receptor TR6 - and corresponding
PT polypeptide, antibody, agonist, antagonist, etc
PS Claim 4; Page 25-27; 34pp; English.
CC This sequence encodes a novel human tumour necrosis factor related
CC receptor, TR6. TR6 polypeptides and polynucleotides can be used in the
CC treatment of chronic and acute inflammation, arthritis, septicaemia,
CC autoimmune diseases (e.g. inflammatory bowel disease, psoriasis),
CC transplant rejection, graft vs. host disease, infection, stroke,
CC ischaemia, acute respiratory disease syndrome, restenosis, brain injury,
CC (acquired autoimmune disease syndrome) AIDS, bone diseases, cancer (e.g.
CC lympho-proliferative disorders), atherosclerosis and Alzheimers disease.
SQ Sequence 3881 BP; 937 A; 922 C; 983 G; 1028 T;

Query Match 7.3%; Score 30.2; DB 1; Length 3881;
Best Local Similarity 59.5%; Pred. No. 10;
Matches 47; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 233 GTCACCTCCACCATCGAGATCTCACACCATGTGTCTCTCACACATGACGACCATG 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3651 GTACATGTCCTTTGATATATATATATATATATATATATATATATATATATATAT 3710

QY 293 AAGCTGTGTCCTTCCTG 311
    ||||| ||||| |||||
Db 3711 TATAGTGTGTCACTCTGCG 3729
```

```
RESULT 10
Q74052
ID Q74052 standard; DNA; 7824 BP.
AC Q74052;
DE 29-JAN-1996 (first entry)
DE Human interleukin-1-beta.
KW Interleukin 1 beta; primer; mRNA; specificity; pharmaceutical; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_difference 1331
    /*tag= a
    /note= "N is undefined"
J07123984-A.
PD 16-MAY-1995.
PF 05-NOV-1993; 275852.
PR 05-NOV-1993; JP-275852.
PA (HITB ) HITACHI CHEM CO LTD.
DR WPI; 95-211627/28.
PT A primer for the detection and the determ. of a specific messenger
PT RNA - can detect and determine specific mRNA(s) with high
PT reliability
PS Example 18; Page 17-20; 35pp; Japanese.
CC Q74052 is interleukin-1-beta cDNA and Q74019-21 are primers used for
CC the amplification of this cDNA. They are used specifically for the
CC detection and isolation of this sequence. The primers have the
CC advantage of high sensitivity and reliability and are useful in the
CC pharmaceutical industry.
SQ Sequence 7824 BP; 2099 A; 1905 C; 1624 G; 2195 T;

Query Match 7.2%; Score 30; DB 1; Length 7824;
Best Local Similarity 55.9%; Pred. No. 15;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 119 ACTTTCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5839 ACTCTACAGCTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 5898

QY 179 CTGGAATTCCTTGAAAGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5899 CACTTATCCAGACACACACACACACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5940

RESULT 11
T23881/C
ID T23881 standard; cDNA to mRNA; 287 BP.
AC T23881;
DE 16-AUG-1996 (first entry)
DE Human gene signature HUMGS05816.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1477; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the-
```


KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN WO9845435-A2.
PD 15-OCT-1998.
PF 10-APR-1998; US-695954.
PR 10-APR-1997; US-835913.
PA (GEMV) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070076/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 210; 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
SQ Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;

Query Match 7.2%; Score 29.8; DB 1; Length 291;
Best Local Similarity 70.2%; Pred. No. 4.7;
Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 115 TATGACCTTGCACACTGACGAGGAGTCTTTCTGACAAATTCCTCTATGAGTC 171
D5 257 TTGTGATTTGACAGAGCTGAGTATTTCTTACACATTTCTCTTAAATC 201

RESULT 15
X20599/c
ID X20599 standard; DNA; 20757 BP.
AC X20599;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI; 99-081273/07.
PT New Isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 661-672; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 20757 BP; 4912 A; 6068 C; 4457 G; 4457 T;
Query Match 7.2%; Score 29.8; DB 1; Length 20757;
Best Local Similarity 49.1%; Pred. No. 26;

Matches 79; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 228 CCAGGCTCACCTCCACCATGCAAGATCTCAACACCATGTTCTGCAACACATGACAGC 287
DB 4855 CCAGATTTCAGCTTCGAAAAAGTGAGCGATGTGCAGCAAGGCAACGGATGGCGATC 4796
QY 288 CATTGAGCCTGTGCTCTTGGCCCGGCTTTTGGGCGGGGATGCAGGAGCGAGGC 347
DB 4795 GCGTGTGACCTGTGGGTACCTTGTCTGCTGGGTCTCCCTTGGGGGAGGTGTGA 4736
QY 348 CCAGACCTGTCTTTACAGAGGCCCCACCCCTCTCTGAGTGGC 388
DB 4735 GTCTCCTGTTGGTGGCTGGACCTAGCTAGCCTGCGAGCGGC 4695
Search completed: March 20, 2000, 09:32:31
Job time: 1114 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2000, 16:07:56 ; Search time 61.26 Seconds
(without alignments)
811.048 Million cell updates/sec

Title: US-09-092-296-7_COPY_5_419

Perfect score: 415

Sequence: 1 ACGGACTTCAGTGTCTCC.....AAATTCGGTATGCTGATT 415

Scoring table: IDENTITY_NUC

Searched: 214294 seqs, 59861208 residues

Database : Issued_Patents_NA:*

Word size : 0

Number of hits that pass the threshold : 428588

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/PTUS9_COMB.seq:*
- 7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.2	10.2	7218	1	US-08-232-463-14
2	30.4	7.3	580	4	US-08-627-610-7
3	30.4	7.3	580	6	PCT-US95-04636-7
4	29.8	7.2	3509	3	US-08-817-436A-1
5	29.2	7.0	1560	1	US-07-669-171-1
6	29.2	7.0	5962	1	US-08-188-582-10
7	29.2	7.0	5962	1	US-08-646-715-10
8	29	7.0	3060	1	US-08-098-141-1
9	29	7.0	1187	1	US-07-709-949-1
10	29	7.0	2259	7	5185254-3
11	28.6	6.9	2662	5	US-08-750-357-8
12	28.4	6.8	6306	1	US-08-466-390-3
13	28.4	6.8	6306	2	US-08-470-950-3
14	28.4	6.8	6306	2	US-08-467-781-3
15	28.4	6.8	6306	2	US-08-195-487-3
16	28.4	6.8	6306	3	US-08-483-924-3
17	28.4	6.8	2262	4	US-08-674-887A-5
18	28.4	6.8	6306	6	PCT-US93-06160-3
19	27.8	6.7	185	3	US-08-332-766A-31
20	27.8	6.7	2782	3	US-08-937-466-1
21	27.8	6.7	2782	4	US-09-172-528-1
22	27.6	6.7	1569	1	US-07-669-171-3
23	27.6	6.7	1853	2	US-08-553-110-2
24	27.6	6.7	5102	2	US-08-494-168-1
25	27.6	6.7	3088	2	US-08-418-444A-1
26	27.6	6.7	3262	3	US-08-678-039A-41
27	27.4	6.6	1364	2	US-08-306-691B-50
28	27.4	6.6	4425	3	US-08-749-169A-1
29	27.4	6.6	2865	3	US-08-749-169A-2
30	27.4	6.6	4425	4	US-09-130-032A-1
31	27.4	6.6	2865	4	US-09-130-032A-2
32	27.4	6.6	3037	4	US-08-938-365-1

c 33	27.4	6.6	1364	6	PCT-US93-06251-65
c 34	27.2	6.6	2043	1	US-07-914-281-1
c 35	27.2	6.6	2043	1	US-08-393-246-1
c 36	27.2	6.6	2043	1	US-08-273-411-4
c 37	27.2	6.6	2121	1	US-08-331-394-3
c 38	27.2	6.6	2121	2	US-08-250-858-3
c 39	27.2	6.6	2121	2	US-08-446-915-3
c 40	27.2	6.6	2121	2	US-08-525-058A-1
c 41	27.2	6.6	2122	2	US-08-485-449-1
c 42	27.2	6.6	2121	3	US-08-744-139-3
c 43	27.2	6.6	2043	4	US-08-696-731-1
c 44	27.2	6.6	2043	6	PCT-US91-00899-1
c 45	27.2	6.6	2121	6	PCT-US95-06639-3

ALIGNMENTS

RESULT 1
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-Fls
: US-08-232-463-14

Query Match 10.2%; Score 42.2; DB 1; Length 7218;
Best Local Similarity 2.8%; Pred. No. 0.0008;
Matches 11; Conservative 214; Mismatches 162; Indels 0; Gaps 0;

QY 4 GGGACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGGCTATGGGCTCGGCC 63

Db 1055 GGAGCTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1114
QY 64 CCTGTGCTCTCTTGACCCICCTTGGCAGCTCAGATGGAACAGGCCGGGTATGACTTT 123
Db 1115 YY 1174
QY 124 GCAACTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTCTATGAGTCCAGCTTCTCGGA 183
Db 1175 YY 1234
QY 184 ATTGCTGAAGAGCTGCTGCTCTCTCTCACTCCCTTCAAGGACACAGCTACCTTCA 243
Db 1235 YY 1294
QY 244 CCATGCAAGATCAACACCATCTTCTGCAACATGACAGCCATTGAAGCTGTGTC 303
Db 1295 YY 1354
QY 304 CTCTTGCCGGGCTTTGGCCGGGATGAGAGGAGCCGCCGACCTGCTTTCA 363
Db 1355 YY 1414
QY 364 GCAGGCCCCACCTCTCTGAGTGGCAA 390
Db 1415 YYYYYYYYYYYYYYYYYYGTACCAA 1441

RESULT 2

US-08-627-610-7/c
; Sequence 7, Application US/08627610
; Patent No. 591997
; GENERAL INFORMATION:
; APPLICANT: Beach, David H.
; APPLICANT: Serrano, Manuel
; APPLICANT: DePinho, Ronald A.
; TITLE OF INVENTION: Transgenic Animals Having Modified Cell-Cycle
; TITLE OF INVENTION: Regulation
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/627,610
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: CSI-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91..480
US-08-627-610-7

Query Match 7.3%; Score 30.4; DB 4; Length 580;
Best Local Similarity 57.3%; Pred. No. 1.4;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 35 GCAGTGGCCACTATGGGCTGCGGCTGCGCCCTTGTCTCTCTTGACCCCTCTTGGCAGC 94
Db 128 GCGGTGGCCAGGCCCGGCTACTGCTGCGCCCAACATGCCCCCTGTCCCGGCTGTGTC 69
QY 95 TCACATGGAACAGGCGCGGATGACTTTGCAACIG 130
Db 68 AGAATGGTCTCTTCGCGCGCGGTGAGATTGCTACAG 33
RESULT 3
PCT-US95-04636-7/c
; Sequence 7, Application PC/TUS9504636
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
; TITLE OF INVENTION: Related Thereto
; NUMBER OF SEQUENCES: 10
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04636
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/346,147
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,511
; FILING DATE: 14-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,812
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,371
; FILING DATE: 14-APR-1994
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91..480
PCT-US95-04636-7

Query Match 7.3%; Score 30.4; DB 6; Length 580;
Best Local Similarity 57.3%; Pred. No. 1.4;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 35 GCAGTGGCCACTATGGGCTGCGGCTGCGCCCTTGTCTCTCTTGACCCCTCTTGGCAGC 94
Db 128 GCGGTGGCCAGGCCCGGCTACTGCTGCGCCCAACATGCCCCCTGTCCCGGCTGTGTC 69
QY 95 TCACATGGAACAGGCGCGGATGACTTTGCAACTG 130
Db 68 AGAATGGTCTCTTCGCGCGCGGTGAGATTGCTACAG 33
RESULT 4
US-08-817-436A-1
; Sequence 1, Application US/08817436A


```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1157 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 62...1015
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 62..119
; PUBLICATION INFORMATION:
; AUTHORS: McLean, J W
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 259
; PAGES: 6498-6504
; DATE: 1984
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1157
US-07-709-949--1

Query Match          7.0%; Score 29; DB 1; Length 1157;
Best Local Similarity 54.1%; Pred. No. 5;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY      6   GACTTCAGTGTCTCTCCATCCCAGGAGCGCACTGCCACACTATGGGGTCTGGGGTGCCCC    65
DB      636  GAGCGGGCCTCAGCGGCATCCGCGCGGCTCGGGCCCTGGGGCCCTGGTGTGAACAGGGCCGGGTG    685

QY      66   TTGTCTCTCTTTGACCCTCTTTGGCAGCTCACATGGAACAAGGCGCGG    114
DB      686  CGGGCGGCACTGTGGGCTCCCTGGCGCGCCAGCGCTACAGGAGCGGG    734

RESULT 10
5185254-3
; Patent No. 5185254
; APPLICANT: LINNENBACH, ALBAN
; TITLE OF INVENTION: GENE FAMILY OF TUMOR-ASSOCIATED ANTIGENS
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/291.583
; FILING DATE: 29-DEC-1988
; SEQ ID NO:3:
; LENGTH: 2259
5185254-3

Query Match          7.0%; Score 29; DB 7; Length 2259;
Best Local Similarity 52.0%; Pred. No. 6.8;
Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY      17   CTCCTCCATCCCAGGAGCGCATGGCCACTATGCGGGTCTGGGGTGGCCCTTGCTCTCTC    76
DB      1120  cgctccacgcgccccatcgccgtcatcgttgttggtcggtggccctcgtgcgccgc    1179

QY      77   TTGACCTCCTTTGGCAGCTCACATGGAACAGGCGGGGTATGACTTTGCACTGAAGCTG    136
DB      1180  atggcgcctcctgggatcaccaaccggagaagtcggggaagtacaagaagtgagatc    1239

QY      137  AAGGA 141
DB      1240  aaggA 1244

RESULT 11

```



```

; Sequence 3, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 33 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILLYA
; AUTHORS: CLEVELAND, DON W
; TITLE: PRIMARY STRUCTURE OF OF NDMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: JOURNAL OF CELL BIOLOGY
; VOLUME: 116
; PAGES: 1395-1408
; DATE: MAR-1992
; IS-08-195-487-3

```

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2000, 08:56:37 ; Search time 538.4 Seconds
(without alignments)
2910.291 Million cell updates/sec

Title: US-09-092-296-7_COPY_5_419
Perfect score: 415
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....AAATTCGGTATGCTGAATT 415

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*

50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
c	1	308.4	74.3	404	61	AI857998	AI857998 w169801.x
c	2	153	36.9	552	82	AQ18761	AQ18761 HS_5511.B
c	3	152.8	36.8	176	21	T94049	T94049 ye33f07.r1
c	4	106	25.5	328	42	AI136523	AI136523 UI-R-C2p-
c	5	46.8	11.3	190	41	AI010074	AI010074 EST204525
c	6	40.8	9.8	201	47	AI535335	AI535335 UI-R-C3-s
c	7	37.4	9.0	1101	79	CNS00K09	AL076960 Drosophil

```
c 8 36.4 8.8 331 31 AA282040
c 9 36.4 8.8 494 34 AA534569
c 10 36.4 8.8 424 38 AA741185
c 11 36.4 8.8 402 40 AA936790
c 12 36.4 8.8 529 41 A1018775
c 13 35.6 8.6 354 70 A143645
c 14 35.2 8.5 537 33 AA17974
c 15 35.2 8.5 460 40 AA909652
c 16 35 35 F06958
c 17 34.8 8.4 451 61 A1829206
c 18 34.8 8.4 939 79 CNS00CNG
c 19 34.4 8.3 378 30 AA243870
c 20 34.4 8.3 238 32 AA376266
c 21 34.4 8.3 715 38 AA746639
c 22 34.4 8.3 329 38 AA769413
c 23 34.4 8.3 412 38 AA769782
c 24 34.4 8.3 458 49 A1632159
c 25 34.4 8.3 470 60 A1803529
c 26 34.2 8.2 509 88 A0838514
c 27 33.6 8.1 429 39 AA862672
c 28 33.4 8.0 258 20 D34884
c 29 33.4 8.0 545 23 T41524
c 30 33.2 8.0 292 40 AA91822
c 31 33.2 8.0 597 63 A0013553
c 32 33 8.0 261 29 AA183985
c 33 32.8 7.9 405 21 R02824
c 34 32.8 7.9 360 35 C43492
c 35 32.8 7.9 240 36 C70784
c 36 32.6 7.9 239 39 AA830716
c 37 32.6 7.9 522 82 A0718755
c 38 32.4 7.8 490 51 A157641
c 39 32.4 7.8 537 64 A078074
c 40 32.4 7.8 523 69 A0128453
c 41 32.2 7.8 339 21 F08745
c 42 32.2 7.8 742 27 Z78408
c 43 32.2 7.8 477 45 A1351147
c 44 32.2 7.8 250 62 A1926123
c 45 32.2 7.8 875 82 A0740265
```

ALIGNMENTS

```
RESULT 1
A1857998/c
LOCUS wj69b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408041 3',
DEFINITION mRNA sequence. EST 26-AUG-1999
ACCESSION A1857998
VERSION A1857998.1 GI:5511614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
Sequencing by: Washington University Genome Sequencing Center
Cloned through: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnlni.gov/bbrp/image/image.html
```

```
Seq primer: -40UP from Gibco
High quality sequence stop: 395.
Location/Qualifiers
source
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/note="organ: lung; Vector: pMT3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pMT3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 97 a 105 c 117 g 84 t 1 others
ORIGIN
Query Match 74.3%; Score 308.4; DB 61; Length 404;
Best Local Similarity 99.4%; Pred. No. 7.2e-79;
Matches 309; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 105 CAGGCGGGGTATGACTTTCGCACTGAAGCTGAAGAGGCTTTTCTGACAAATTCCTCT 164
DB 311 CAGGCGGGGTATGACTTTCGCACTGAAGCTGAAGAGGCTTTTCTGACAAATTCCTCT 252
QY 165 ATGAGTCACAGCTTCCTGGAATTCCTGAAAGCTTCGCTCCTCCTCCATCCCTTCAG 224
DB 251 ATGAGTCACAGCTTCCTGGAATTCCTGAAAGCTTCGCTCCTCCTCCTCCTTCAG 192
QY 225 GGACGAGGCTCACCTCCACCATCAAGATCTCAACACCATGTTGCTGCAACACATGAC 284
DB 191 GGACGAGGCTCACCTCCACCATCAAGATCTCAACACCATGTTGCTGCAACACATGAC 132
QY 285 AGCCATTGAGCTGTGCTCTTTCGCGCGGGCTTTTGGCGCGGGATCGCAGGAGGCGAG 344
DB 131 AGCCATTGAGCTGTGCTCTTTCGCGCGGGCTTTTGGCGCGGGATCGCAGGAGGCGAG 72
QY 345 GCCCGACCTGTCTTTCAGCAGGCGCGCGCGCTCCTGAGTGGAATAAATAAATTCGG 404
DB 71 GCCCGACCTGTCTTTCAGCAGGCGCGCGCGCTCCTGAGTGGAATAAATAAATTCGG 12
QY 405 TATGCTGAAT 415
DB 11 TATGCTGAAT 1
RESULT 2
A0718761/c
LOCUS A0718761 552 bp DNA GSS 13-JUL-1999
DEFINITION HS_5511_B2_F09_77A RPII-11 Human Male BAC Library Homo sapiens
genomic clone Plate-1087 Col-18 Row=L, genomic survey sequence.
ACCESSION A0718761
VERSION A0718761.1 GI:5468077
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
```


451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics

Seq primer: M13 Forward.

FEATURES
 source
 1. 328
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C2p-nq-e-02-0-UI"
 /clone_lib="UI-R-C2p"
 /dev_stage="adult"
 /lab_host="DRI08 (Life Technologies)"
 /note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

BASE COUNT 62 a 77 c 98 g 91 t

Query Match 25.5%; Score 106; DB 42; Length 328;
Best Local Similarity 66.8%; Pred. No. 1.8e-20;
Matches 151; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 90 GCAGCTCAGATGACAGCGCGGGTATGACCTTCGCACTGAGCTGAGGAGTCTTTTC 149
 Db 327 GCAGCTCAGTCCGAGGCGGCACAAAGGTGACCTTCGCAAGTAAAGCTGACCGAGGATCTC 268

QY 150 TGACAAATTCCTCTATGAGTCCAGCTTCCTGGAATTCCTGGAAGCTCTGCTCTCTCC 209
 Db 267 AGGCCAAGACCTCCCAAGACTCCGGCTTTCTGGACATGCTCCAAAGATCTGCTCTCTCC 208

QY 210 TCCATCTCCCTCAGGACGAGCTCACCCTCCAGCATGCAAGATCTCAACCATGTTG 269
 Db 207 TCCACCTCTCAGGGGACCAATGTCACCTTCATCATAAAGGGCCACACACACCTTA 148

QY 270 TCTGCACACATGACAGCCATTGAAGCCTGTGCTCTTCTTGGCCCG 315
 Db 147 CTTGCAGAGCTGAGAGCAGCTGAAGCCTGCGTGGGCGCTGGAGCGG 102

RESULT 5
AI010074/c
LOCUS AI010074 190 bp mRNA EST 15-JUN-1998
DEFINITION EST204525 Normalized rat lung, Bonto Soares Rattus sp. cDNA clone RIJBT38 3' end, mRNA sequence.
ACCESSION AI010074

VERSION AI010074.1 GI:3223906
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 190)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
JOURNAL Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
COMMENT Unpublished (1998)
 On Jan 19, 1998 this sequence version replaced gi:2150555.
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES
 source
 1. 190
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RIJBT38"
 /clone_lib="Normalized rat lung, Bonto Soares"
 /note="Organ: lung; Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 39 a 51 c 59 g 41 t

Query Match 11.3%; Score 46.8; DB 41; Length 190;
Best Local Similarity 65.1%; Pred. No. 0.0019;
Matches 69; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 210 TCCATCTCCCTTCAGGACGACGCTCCACCTCCACCATGCAAGATCTCAACACCATGTTG 269
 Db 190 TCCACTCTCACCAGGACCAATGTCACCTTCATCATAAAGGGCCACACACACCTTA 131

QY 270 TCTGCACACATGACACCATGTAAGCCTGTGCTCTTCTTGGCCCG 315
 Db 130 CTTGCAGAGCCTGAGAGCAGCTGAGCCTGCTGGGCGCTGGAGCGG 85

RESULT 6
AI535335/c
LOCUS AI535335 201 bp mRNA EST 18-MAR-1999
DEFINITION UI-R-C3-sw-e-12-0-UI.s1 UI-R-C3 Rattus norvegicus cDNA clone
 UI-R-C3-sw-e-12-0-UI 3', mRNA sequence.
ACCESSION AI535335
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 201)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT On May 18, 1998 this sequence version replaced gi:3136861.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone Distribution: clones will be available through Research Genetics (www.resgen.com). Seq primer: M13 Forward.

collaboration with the Berkeley Drosophila genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCR-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

D	b	922	TBGSYGBGNGSYVCGSGSGSRSTCSCTBAKHSYCCSMCMGCHVTMSSTKH	863
			: :	
Q	y	242	CACCATCGAAGATCTCAACACCATTGTGTCTGCAACACATCAGAGCATTTGAAGCCTGTG	301
			: :	
D	b	862	CVCCTVTWBNEDTTCACVCMCTTTTTTTTTTTTAAAMVRNTRKNVSMACTDMWRVKGYG	803
			: :	
Q	y	302	TCCTTTCTGGCCCCGGGCTTTTGCGCCGGGATCGAGAGGAGCGGCCCCACCTGTCTTT	361
			: :	
D	b	802	TDTTGKGSYGGEKTDGKKKGKGSSSGGCKKKKTGVBGRBSTKYIYMSCNTKMGTVYWHDD	743
			: :	
O	v	362	CAGCAGSCCCCACCC	377

DB	742	TTVCCCAACCAACCC	727	
RESULT	8			
AA282040/c				
LOCUS	AA282040	331 bp	mRNA	EST 14-AUG-1997
DEFINITION	zs89a03.s1 NCI_CGAP_GCBI Homo sapiens cDNA clone IMAGE:704620 3',			
	mRNA sequence.			
ACCESSION	AA282040			
VERSION	AA282040			
KEYWORDS	AA282040.1 GI:1924872			
SOURCE	EST.			
	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
	Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 331)			
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .			
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).			

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1402231.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LILN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 891 Std Error: 0.00
Seq Primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 242.
Location/Qualifiers
1. .331
/organism="Homo sapiens"
/db_xref="GDB:5854344"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_C03"
/clone="IMAGE:704620"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-).
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(GT) primer
[5'-TGTTACCAATCTGAATGCGGCGCGCTCAATTTTTCCTTTTTCCTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES
source

BASE COUNT 97 a 61 c 51 g 122 t
ORIGIN
Query Match 8.8%; Score 36.4; DB 31; Length 331;
Best Local Similarity 56.8%; Pred. No. 2.3;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Oy 133 GCTGAAGAGCTTTCTGACAAATTCCTCTATGAGTCAGCTTCCTGGAATGCTTGA 192
|||||
Db 318 GATGATGACTACTGCTATATATAATCCACTAAATCCAGCTACCAAGAACTGCCTGG 259
||
Oy 193 AAAGCTCTCCCTCCCTCCATCTCCCTTCAGGACCAAGCTCACCCTCCACCATGCA 250
||
Db 258 AACTGTGCCATGCAATTTTCTTTTAAAGACCAAGTGTGATAGTAGGCCATGCA 201
||
RESULT 9
AA534569/c 494 bp mRNA EST 21-AUG-1997
LOCUS n177c11.s1 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:925940 3',
DEFINITION mRNA sequence.
ACCESSION AA534569
VERSION AA534569.1 GI:2278822
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802984.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 370.
Location/Qualifiers
1. .494

FEATURES
source

BASE COUNT 139 a 108 c 94 g 153 t
ORIGIN
Query Match 8.8%; Score 36.4; DB 34; Length 494;
Best Local Similarity 56.8%; Pred. No. 2.6;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Oy 133 GCTGAAGAGCTTTCTGACAAATTCCTCTATGAGTCAGCTTCCTGGAATGCTTGA 192
|||||
Db 296 GATGATGACTACTGCTATATATAATCCACTAAATCCAGCTACCAAGAACTGCCTGG 237
||
Oy 193 AAAGCTCTCCCTCCCTCCATCTCCCTTCAGGACCAAGCTCACCCTCCACCATGCA 250
||
Db 236 AACTGTGCCATGCAATTTTCTTTTAAAGACCAAGTGTGATAGTAGGCCATGCA 179
||
RESULT 10
AA741185/c 424 bp mRNA EST 08-FEB-1998
LOCUS ob30a02.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1325162 3',
DEFINITION mRNA sequence.
ACCESSION AA741185
VERSION AA741185.1 GI:2779777
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 5, 1998 this sequence version replaced gi:2581148.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
unknown library type
Insert Length: 742 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 387.
Location/Qualifiers
1. .424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1325162"


```

BASE COUNT      146 a 114 c 104 g 165 t
ORIGIN

Query Match      8.8%; Score 36.4; DB 41; Length 529;
Best Local Similarity 56.8%; Pred. No. 2.7;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 133 GGTGAAGAGTCTTTTTCACAAATCTCTATGAGTCCAGCTTCTGGAATGCTTGA 192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 GATGATGCTACTGCTATATATATATATATATATATATATATATATATATATAT 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 193 AAAGCTGCTGCTCTCTCCATCTCCCTTCAGGACCAAGCTCACCCTCCACCAGCA 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 AACTGTGCCATGCAITTTTTTTTTTTTAAAGACCAAGTGTGATAGTGGCCATGCA 182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AW143645/c      354 bp mRNA EST 30-OCT-1999
LOCUS      RGTB23941 Normalized rat embryo, Bonto Soares Rattus sp. cDNA clone
DEFINITION
ACCESSION      AW143645
VERSION
KEYWORDS
SOURCE
ORGANISM      Rattus sp.
REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
              1 (bases 1 to 354)
              Lee N.H., Glodek A., Chandra, I., Mason, T.M., Quackenbush, J.,
              Korlavage, A.R. and Adams, M.D.
              Rat Genome Project: Generation of a Rat EST (RAT) Catalog & Rat
              Gene Index
              Unpublished (1998)
              On Apr 30, 1999 this sequence version replaced gi:4727610.
              Contact: Lee, NH
              ATCC
              The Institute for Genomic Research
              9712, Medical Center Drive, Rockville, MD 20850, USA
              Tel: (301)-838-3529
              Fax: (301)-838-0208
              Email: nhlee@tigr.org
              For clone availability, additional sequence and expression
              information related to this EST please check the TIGR Rat Gene
              Index (http://www.tigr.org/tdb/rat/rat.html). To order a clone
              contact the ATCC (http://www.atcc.org/atcc.html).
              Seq primer: M13 Reverse.
              Location/Qualifiers
                1..354
                /organism="Rattus sp."
                /db_xref="taxon:10118"
                /clone="RG14X27"
                /clone_lib="Normalized rat embryo, Bonto Soares"
                /dev_stage="embryo 8, 12, 18 dpc"
                /note="Vector: pT73pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      106 a 76 c 82 g 90 t
ORIGIN

Query Match      8.6%; Score 35.6; DB 70; Length 354;
Best Local Similarity 45.9%; Pred. No. 4;
Matches 122; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 118 GACTTTGCACTGAAGCTGAAGAGTCTTTTCGACAAATTCCTCCATATGATCCAGCTT 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GATTTTCTCTCTTAACTGCATCTACTTTCGAGAAAGATCTTCAAGTGTGCTGGAGCT 210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 178 CTGTGAATGTGTTGAAGCTCTGCTCCCTCCATCTCCCTTCAGGACAGCGTCAAC 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 CCACACCTGATATATGATCGAGTCTTTTCTCCAGTTCACATGATTTTCAGCATCA 150
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 238 CCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACATGACAGCATTAAGCC 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 149 CTGGTCCATGTCACATTCATCTTGGGAACAGACGTTTCAGTCCCGAATGCAAC 90
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 TGTGTCCTCTTGGCCCGGGCTTTTGGCCGGGATGACAGGAGCGAGCCCGACCTGT 357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 TCTTTAATGAATGGTCTTGGCTTTGAGCGCGCGGAAAGCGGATAGTAGACGCT 30
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 CTTTTCAGCAGGCGCCCGACCCCTCTCTGA 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 CAGCGCGGCTCTCCACCACTGA 4
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
AA417974      537 bp mRNA EST 16-OCT-1997
LOCUS      zv97c08.s1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:767726 3'
DEFINITION      similar to WP:F59B2.3 CE00231 N'-ACETYL-GLUCOSAMINE-6-PHOSPHATE
              DEACETYLASE ; mRNA sequence.
ACCESSION      AA417974
VERSION
KEYWORDS
SOURCE
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 537)
              Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
              Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
              Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
              White, Y., Wyllie, T., Waterston, R. and Wilson, R.
              WashU-Merck EST Project 1997
              Unpublished (1997)
              On Sep 12, 1996 this sequence version replaced gi:1393089.
              Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@wustl.wustl.edu
              This clone is available royalty-free through LNL; contact the
              IMAGE Consortium (infoimage.llnl.gov) for further information.
              Possible reversed clone: similarity on wrong strand
              Seg primer: -41m3 fwd. ET from Amersham
              High quality sequence stop: 502.
              Location/Qualifiers
                1..537
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:767726"
                /clone_lib="Soares_NhMPu_S1"
                /tissue_type="Pooled human melanocyte, fetal heart, and
                pregnant uterus"
                /lab_host="DH10B"
                /note="Organ: mixed (see below); Vector: pT73D-Pac
                (Pharmacia) with a modified polylinker; Site_1: Not I;
                Site_2: Eco RI. Equal amounts of plasmid DNA from three
                normalized libraries (melanocyte 2NDHM, pregnant uterus
                NDHPT, and fetal heart NDH19W) were mixed, and ss circles
                were made in vitro. Following HAP purification, this DNA
                was used as tracer in a subtractive hybridization
                reaction. The driver was PCR-amplified cDNAs from pools of
                5,000 clones made from the same 3 libraries. The pools
                consisted of I.M.A.G.E. clones 260232-265223,
                340488-345479, and 484488-489479."
BASE COUNT      90 a 193 c 166 g 88 t
ORIGIN

```

FEATURES

source

Query Match

Best Local Similarity 47.8%; Pred. No. 6.1;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2000, 19:58:50 ; Search time 360.16 Seconds
(without alignments)
-1972.773 Million cell updates/sec

Title: US-09-092-296-7_COPY_51_284
Perfect score: 234
Sequence: 1 ATGGGGTCTGGGTGCCCT.....ACCATGTTGTCGCAACACA 234
Scoring table: IDENTITY_NUC
Searched: 821193 seqs, -1518192014 residues
Database: GenEmbl.*
Word size: 0
Number of hits that pass the threshold: 1642386
1: gb_ba1.*
2: gb_ba2.*
3: gb_em.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pi1.*
8: gb_pi2.*
9: gb_pi3.*
10: gb_pi4.*
11: gb_pi5.*
12: gb_pi6.*
13: gb_pi7.*
14: gb_pi8.*
15: gb_pi9.*
16: gb_pi10.*
17: gb_pi11.*
18: gb_pi12.*
19: gb_pi13.*
20: gb_pi14.*
21: gb_pi15.*
22: gb_pi16.*
23: gb_pi17.*
24: gb_pi18.*
25: gb_pi19.*
26: gb_pi20.*
27: gb_pi21.*
28: gb_pi22.*
29: gb_pi23.*
30: gb_pi24.*
31: gb_pi25.*
32: gb_pi26.*
33: gb_pi27.*
34: gb_pi28.*
35: gb_pi29.*
36: gb_pi30.*
37: gb_pi31.*
38: gb_pi32.*
39: gb_pi33.*
40: gb_pi34.*
41: gb_pi35.*
42: gb_pi36.*
43: gb_pi37.*
44: gb_pi38.*
45: gb_pi39.*
46: gb_pi40.*
47: gb_pi41.*
48: gb_pi42.*
49: gb_pi43.*

50: gb_pi13.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	92.4	39.5	192650	10	AB023048	Homo sapi
2	92.4	39.5	200000	10	AP000511	AP000511 Homo sapi
3	92.4	39.5	47323	11	AC005937	AC005937 Homo sapi
C 4	36.4	15.6	118276	11	AC004148	AC004148 Homo sapi
C 5	35.4	15.1	207984	33	AC006329	AC006329 Homo sapi
C 6	34.2	14.6	3626	12	D86631	D86631 Mus sp. DNA
C 7	33.4	14.3	65673	42	AC012184	AC012184 Homo sapi
C 8	33.2	14.2	141762	10	HS198C21	AL034400 Human DNA
C 9	33	14.1	116218	11	AC005090	AC005090 Homo sapi
C 10	32.8	14.0	30452	10	AB014077	AB014077 Homo sapi
C 11	32.8	14.0	100000	10	AP000513	AP000513 Homo sapi
C 12	32.8	14.0	43600	11	AC004209	AC004209 Homo sapi
C 13	32.8	14.0	159515	12	AC002327	AC002327 Mus muscu
C 14	32.6	13.9	172048	10	HS179M16	Z95152 Homo sapien
C 15	32.4	13.8	184437	16	HRVU20824	U20824 Equine hepr
C 16	32.2	13.8	2122	3	PIGCHAT	J03021 Porcine cho
C 17	32.2	13.8	2122	3	PIGCHAT	M27736 Pig choline
C 18	32.2	13.8	133203	11	AC003036	AC003036 Homo sapi
C 19	32.2	13.8	36676	11	HS7858B16	AL056768 Human DNA
C 20	32.2	13.8	167228	40	AC005552	AC005552 Homo sapi
C 21	32	13.7	137615	42	AC011736	AC011736 Homo sapi
C 22	31.8	13.6	97580	11	HS591M18	AL031594 Human DNA
C 23	31.6	13.5	45632	35	CEZC3527	AF067216 Caenorhab
C 24	31.6	13.5	2803	2	TN191334	U93334 Thermotoga
C 25	31.6	13.5	140856	10	NC002430	AC002430 Human BAC
C 26	31.6	13.5	170842	40	AC007384	AC007384 Homo sapi
C 27	31.2	13.3	17324	7	AF001064	AF001064 Archaeogl
C 28	31.2	13.3	5419	7	PTC3BP	X61915 P. thuberyi
C 29	31	13.2	4824	9	AB014540	AB014540 Homo sapi
C 30	31	13.2	133457	11	AC003999	AC003999 Human PAC
C 31	31	13.2	168196	11	CNS01DTS	AL112857 Human chr
C 32	31	13.2	6457	11	HS953387	HS953387 Homo sapien
C 33	31	13.2	126844	33	AL153347	AL153347 Homo sapi
C 34	31	13.2	45123	40	HSARBYFL7	U72843 Homo sapien
C 35	31	13.2	183728	41	AC009532	AC009532 Homo sapi
C 36	31	13.2	183728	41	AC009538	AC009538 Homo sapi
C 37	31	13.2	97358	44	AC011039	AC011039 Homo sapi
C 38	30.8	13.2	7218	5	166494	166494 Sequence 14
C 39	30.8	13.2	1004	7	SOYHRGPA	L20229 Soybean hyd
C 40	30.8	13.2	116217	8	AFAC005955	AC005955 Arabidops
C 41	30.8	13.2	4434	8	GN044938	U44938 Glycine max
C 42	30.8	13.2	5280	9	AB011114	AB011114 Homo sapi
C 43	30.8	13.2	90347	11	HS465J14	Z84477 Human DNA s
C 44	30.8	13.2	2141	12	AF073309	AF073309 Mus muscu
C 45	30.8	13.2	182123	33	AC002399	AC002399 Homo sapi

ALIGNMENTS

RESULT 1
AB023048 192650 bp DNA 20-NOV-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:53J9, complete sequence.
ACCESSION AB023048
VERSION AB023048.1 GI:5672603
KEYWORDS HTG.
SOURCE Homo sapiens cell_line:978SK DNA, clone:53L9.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo

REFERENCE
AUTHORS

1. (sites)
Shiina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T.,
Kikawa, E., Taka, K., Tomizawa, K., Okueki, N., Kuvano, Y.,
Wakabe, K., Fukumori, Y., Itakura, S., Sugawara, C., Ono, A.,
Yamazaki, M., Tashiro, H., Ando, A., Ikemura, T., Soeda, E., Kimura, M.,
Nairan, S., and Inoko, H.
Molecular dynamics of MHC genes is unraveled by sequence analysis of
the 1,736,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
2. (bases 1 to 192650)
Shiina, T. and Takishima, N.
Direct Submission
Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shiina, Tokai University School of Medicine, Department of
Molecular Life Science 2, Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail: tashina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)

FEATURES
source

Location/Qualifiers
1. 192650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="9785g"
/chromosome="6"
/clone="53L9"
/map="6p21.3"
BASE COUNT 49862 a 44743 c 45833 g 52212 t
ORIGIN

Query Match 39.5%; Score 92.4; DB 10; Length 192650;
Best Local Similarity 98.9%; Pred. No. 4e-18;
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 59 CAGGCCGGGTATGACTTGCACCTGAAGCTGAGGAGTCTTTTCTGACAAATTCCTCT 118
DB 177979 CAGGCCGGGTATGACTTGCACCTGAAGCTGAGGAGTCTTTTCTGACAAATTCCTCT 178038
QY 119 ATGAGTCCAGCTTCTGGAATGCTTGAAGAAGCT 152
DB 178039 ATGAGTCCAGCTTCTGGAATGCTTGAAGAAGT 178072

RESULT 2
AP000511

LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
DEFINITION section 10/20.
ACCESSION AP000511
VERSION AP000511
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

Shiina, S., Tamiya, G., Oka, A., and Inoko, H.
Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
Published Only in DataBase (1999) in press
2. (bases 1 to 200000)
Hirakawa, M., Yamaguchi, H., Imai, K. and Shimada, J.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Database Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail: mika@ast.tokyo.jst.go.jp,
URL: http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)

COMMENT

This sequence is conducted by Tokai University as a JST sequencing
team
Principal Investigator: Hidetoshi Inoko Ph.D
Phone: +81-463-93-1121 Fax: +81-463-94-8884
The sequence is submitted by Human Genome Sequencing in ALIS

Project of JST
Japan Science and Technology Corporation (JST)
5-3 Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www-alis.tokyo.jst.go.jp.

FEATURES
source

Location/Qualifiers
1. 200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
108774..108864
/standard_name="D6S2088"
/note="SHGC-12985; The location is between each flanking
site of PCR primers."
/db_xref="GDB:735288"
113648..115906
/standard_name="D6S1898"
/note="WI-9418; The location is between each flanking site
of PCR primers."
/db_xref="GDB:678272"
complement(115665)..115896
/note="D6A0VH10; The location is between each flanking site
of PCR primers."
/db_xref="GDB:443382"
142030..142159
/standard_name="D6S2203"
/note="SHGC-16870; The location is between each flanking
site of PCR primers."
/db_xref="GDB:741215"
complement(142085)..142322
/note="RR1813; The location is between each flanking site
of PCR primers."
/db_xref="GDB:4573021"
complement(142479)..142740
/standard_name="D6S1851"
/note="SHGC-10808; The location is between each flanking
site of PCR primers."
/db_xref="GDB:675281"
complement(164525)..164665
/note="SHGC-3064; The location is between each flanking
site of PCR primers."
/db_xref="GDB:1234116"
complement(169209)..169574
/standard_name="D6S952"
/note="UT5233; The location is between each flanking site
of PCR primers."
/db_xref="GDB:313481"
BASE COUNT 52605 a 47531 c 49366 g 50498 t
ORIGIN

Query Match 39.5%; Score 92.4; DB 10; Length 200000;
Best Local Similarity 98.9%; Pred. No. 4e-18;
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 59 CAGGCCGGGTATGACTTGCACCTGAAGCTGAGGAGTCTTTTCTGACAAATTCCTCT 118
DB 110306 CAGGCCGGGTATGACTTGCACCTGAAGCTGAGGAGTCTTTTCTGACAAATTCCTCT 110365
QY 119 ATGAGTCCAGCTTCTGGAATGCTTGAAGAAGCT 152
DB 110366 ATGAGTCCAGCTTCTGGAATGCTTGAAGAAGT 110399

RESULT 3
AC005937

LOCUS Homo sapiens clone UNGC:370M23.002 from 6p21, complete sequence.
DEFINITION AC005937
ACCESSION AC005937
VERSION AC005937.1
KEYWORDS HG.


```
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 59 CAGGCGGGTATGACCTTTGCAACTGAAGTGAAGGAGTCTTTTCTGACAAATTCCTCT 118
|||||
Db 35465 CAGGCGGGTATGACCTTTGCAACTGAAGTGAAGGAGTCTTTTCTGACAAATTCCTCT 35524
|||||

QY 119 ATGAGTCCAGCTCCCTGGATGCTTGCARAAGCT 152
|||||
Db 35525 ATGAGTCCAGCTCCCTGGATGCTTGCARAAGGT 35558
|||||

RESULT 4
AC004148/c
LOCUS AC004148.1
DEFINITION Homo sapiens chromosome 17, clone HCV524C5, complete sequence.
ACCESSION AC004148
KEYWORDS AC004148.1 GI:3482960
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
Notes:
1. (bases 1 to 118276)
Birren,B., Fasmann,K., McKernan,K., Nusbaum,C., Richardson,P.,
Boutwell,C., Byrne,S., Cantu,C., Castle,A., Ceny,J., Cooke,E.,
Daly,M.J., Depayre,E., Devon,K., Dewar,K., Donelan,L., Dorette,B.,
Stemmler,S., Geraghty,K., Gilman,T., Gray,D., Gage,B., Gardyna,S.,
Gehrmann,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Linton,L.,
Mackenzie,J., Marquis,N., McEwan,P., McGurk,A., Melgrim,J.,
Molla,M., Morris,W., Morrow,J., Nacman,A., Naylor,J., O'Connor,T.,
Pavlin,B., Peterson,K., Ranganathan,S., Riley,K., Roberts,D.,
Rollins,G., Rossello,R., Roy,A., Shyam,K., Soohoo,S.,
Stange-Thomann,N., Stilwell,J., Stone,C., Strickland,C., Sydney,K.,
Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y.,
Ye,W.J., Zentgraf,I., Zhao,J. and Zody,M.
Direct Submission
Submitted (15-FEB-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 118276)
Birren,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,M., Baker,J., Baldwin,J., Barna,N., Becklerly,R., Bend,J.,
Boutwell,C., Brown,A., Castle,A., Ceny,J., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W.,
Forrest,C., Funke,R., Gage,B., Gardyna,S., Geraghty,K., Grant,G.,
Hagos,B., Heaford,A., Herens,L., Horton,L., Howland,J.C.,
Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Melgrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nahf,R., Naylor,J., Nalloff,M., O'Connor,T., O'Donnell,P.,
Pavlin,B., Peterson,K., Riley,K., Roberts,D., Roy,A.,
Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (28-AUG-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 28, 1998 this sequence version replaced gi:3402743.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp-genome.washington.edu/RM/RepeatMasker.html.
location/Qualifiers
1. 118276
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HCV524C5"

TITLE
JOURNAL
REFERENCE
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
```

```
/clone_lib="Research Genetics/Cal Tech CTRB-HSP-C (plates
195-384)"
/map="17"
/chromosome="17"
/complement(295..592)
/rpt_family="AluSx"
/complement(1063..1204)
/rpt_family="L1MB8"
1229..1276
/rpt_family="(CA)n"
/complement(1279..1468)
/rpt_family="L1MB8"
/complement(1612..1852)
/rpt_family="AluSx"
/complement(1975..2142)
/rpt_family="AluSx"
/complement(2143..2155)
/rpt_family="AluSg"
/complement(2156..2194)
/rpt_family="(TAAA)n"
/complement(2195..2480)
/rpt_family="AluSg"
/complement(2481..2615)
/rpt_family="AluSx"
/complement(2635..2662)
/rpt_family="POLY_A"
2949..3260
/rpt_family="AluJb"
/complement(3391..3401)
/rpt_family="AluSg"
/complement(3402..3428)
/rpt_family="(TAAA)n"
/complement(3439..3677)
/rpt_family="AluSg"
3714..3884
/rpt_family="FRAM"
4143..4219
/rpt_family="L2"
/complement(5050..5364)
/rpt_family="AluSg"
6907..7148
/rpt_family="L2"
7181..7428
/rpt_family="L2"
7508..7647
/rpt_family="L2"
/complement(9623..9918)
/rpt_family="AluSx"
10006..10304AluSg1"
/rpt_family="AluSg"
10312..10476AluSg"
10508..10530
/rpt_family="AluSg"
/rpt_family="AluSg"
/complement(15532..15839)
/rpt_family="AluSx"
/complement(15844..16126)
/rpt_family="AluSx"
17035..17334
/rpt_family="AluSx"
17423..17720
/rpt_family="AluSx"
/complement(18239..18513)
/rpt_family="AluJb"
18587..18814
/rpt_family="MIR"
/complement(20103..20260)
/rpt_family="L1MB3A"
/complement(20340..20494)
/rpt_family="AluJb"
/complement(20495..20793)
/rpt_family="AluSx"
/complement(20794..20934)
```


114	CTCCTGATGAGTCAGCTTCCTGGAATTCGCTGAAAGAGCTCTGCTCTCTCTCGATCTCC	173
QY		
5004	CTTTTCCACAAAGACATCCCGGCCAGCCGAGCAGACCCCTTCGTGGGCCCTTATCTCGT	4945
DB		
174	TTCCAGGGACAGCGTCACCTCCACCATGTCAGAGATCTCAACACCATCTGTGTGTCAGAC	233
QY		
4944	CTGAGCGCCCTGGGAATCCACCCACACAGAATTTCTATACCTCTGTTCTCTCAAGC	4885
DB		
234	A 234	
OY		
4884	A 4884	
DB		

RESULT	6
D86631/c	DNA
LOCUS	3626 bp
DEFINITION	Mus sp. DNA for oxytocin receptor, partial cds.
ACCESSION	D86631
VERSION	G1:1902965
KEYWORDS	oxytocin receptor.
SOURCE	Mus sp. (strain:129SV) ES cell DNA.
ORGANISM	Mus sp.
REFERENCE	Eukariyota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1. (bases 1 to 3626)
TITLE	Kubota, Y.
JOURNAL	Direct Submission
DATE	Submitted (26-JUL-1996) to the DDBJ/EMBL/Genbank databases. Yasue Kubota, Osaka University Medical School, Department of Obstetrics and Gynecology; 2-2 Yamadaoka, Suita, Osaka 565, Japan. (E-mail:kubotay@yume.med.osaka-u.ac.jp, Tel:81-6-879-3351, Fax:81-6-879-3359)
REFERENCE	2. (bases 1 to 3626)
AUTHORS	Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Saij, F. and Murata, Y.
TITLE	Structure and Expression of the Mouse Oxytocin Receptor Gene Unpublished (1996)
REFERENCE	3 (sites)
AUTHORS	Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Saij, F. and Murata, Y.
TITLE	Structure and expression of the mouse oxytocin receptor gene
JOURNAL	Mol. Cell. Endocrinol. 124 (1-2), 25-32 (1996)
MEDLINE	97179034

```

location/Qualifiers
1..3626
/organism="Mus sp."
/strain="129SV"
/db_xref="taxon:10095"
/cell_type="ES cell"
1..1530
/promoter
exon
1531..1798
/number=1
1799..2240
/intron
/number=1
2241..2393
/exon
/number=2
2394..2485
/intron
/number=2
2486..3547
/exon
/number=3
2629..3546
CDS
/codon_start=1
/product="oxytocin receptor"
/protein_id="BAAL8996.1"
/db_xref="GI:1902966"
/translation="MSGTPAANWGSITELDGSVPPEAGNLTAGPPRNEALARVEA
VLCLILFLALSGNACVLLAARTTRKHHSRLFFFMKHSIADLVAVFQVLPOLLNDIT
FRYPGDALCRVLYQVYGFASYTLLILSLDRLCAICQPLSRRTDRLLAVDT
WGLCVASVPQVHIFSLREVDGFDCWAFIQPMGKAVYTWITLAVIIVPVILAA
CYGLISFKTIWNLRLKATAAAAAAEGSDAAGGAAALARSVSKLISKARITKVT
PIVLAFIVCTITPEFFVQWVSVDVNPAPK"
3548..3626
/intron
/number=3

```

	BASE COUNT	782 a	921 c	1131 g	792 t	
	ORIGIN					
	Query Match		14.6%	Score 34.2;	DB 12;	Length 3626;
	Best Local Similarity		60.0%	Pred. NO. 2.8;		
	Matches 57;		Conservative 0;	Mismatches 38;	Indels 0;	Gaps 0;
QY	106	ACAAATCTCTCTATGATGACCTCTCGAATGCTTGAAGCTGTGCGTCCCTCC	165			
Db	2085	AAAAGCTTTTACTAGGGGTACATCTTCCTGGGTCCCCCAAGGCCATCCTCTTCAC	2036			
QY	166	CATCTCCCTTTCAGGACACGCTCACCTCCACCA	200			
Db	2025	CTCCTGCTTCCAGACTCAGCCCGACGCTCCACAA	1991			

```

RESULT 7
ACOL1184/c
LOCUS
DEFINITION Homo sapiens chromosome 16 clone RPCR-11_529k1, *** SEQUENCING IN
PROGRESS ***, 40 unordered pieces.
ACOL1184
ACOL1184.1 GI:6011684
VERSION
KEYWORDS
SOURCE HTG; HTGS_PHASE1.
human
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 65673)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
1 (bases 1 to 65673)
DOE Joint Genome Institute.
2 (bases 1 to 65673)
DOE Joint Genome Institute.
Direct Submission (21-OCT-1999) Production Sequencing Facility, DOE Joint
Submitted (21-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.lgi.doe.gov.
COMMENT

```

* www.jci.doe.gov. * working draft' sequence. It currently
* ADLIS: this is * contains 0 contigs. The true order of the places
* cannot be known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* * be preserved.	760: contig of 760 bp in length
* * 1	gap of unknown length
* * 761	1560: contig of 800 bp in length
* * 1561	gap of unknown length
* * 2303	2303: contig of 743 bp in length
* * 2304	gap of unknown length
* * 3255	3255: contig of 952 bp in length
* * 3256	gap of unknown length
* * 4246	4246: contig of 991 bp in length
* * 4247	gap of unknown length
* * 5603	5603: contig of 1357 bp in length
* * 5604	gap of unknown length
* * 6762	6762: contig of 1159 bp in length
* * 6763	gap of unknown length
* * 7004	7004: contig of 242 bp in length
* * 7005	gap of unknown length
* * 8046	8046: contig of 1042 bp in length
* * 8047	gap of unknown length
* * 9309	9309: contig of 1263 bp in length
* * 9310	gap of unknown length
* * 10988	10988: contig of 1679 bp in length
* * 10989	gap of unknown length
* * 12037	12037: contig of 1049 bp in length
* * 12038	gap of unknown length
* * 13400	13400: contig of 1363 bp in length
* * 13401	gap of unknown length

Db 39755 TAAACTTCAAGCGCAAAATTCCTGTGATGCA 39725

RESULT	8				
HS198C21	141762 bp	DNA	PRI	23-NOV-1999	
LOCUS	Human DNA sequence from clone 198C21 on chromosome Xq26.1-26.3				
DEFINITION	Contains gpC4 (glypican 4), ESTs, STss and GSS, complete sequence.				
ACCESSION	AL034400				
VERSION	AL034400.2	GI:4455461			
KEYWORDS	HTG; GP4.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Heath,P.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NAR-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk				
COMMENT	<p>On Mar 21, 1999 this sequence version replaced gi:3980345.</p> <p>Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.</p> <p>IMPORTANT: This sequence is not the entire insert of clone 198C21. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.</p> <p>The true right end of clone DJ358H7 is at 100 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known, annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.</p> <p>This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX</p> <p>198C21 is from the library RPC16 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR.pPAC4.</p>				
FEATURES	Location/Qualifiers				
source	1..141762				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="X"				
	/map="q26.1-26.3"				
	/clone="rp6-198C21"				
	/clone_lib="RPC1-6"				
repeat_region	1..337				
	/note="MULTI repeat: matches 100..500 of consensus"				
repeat_region	737..810				
	/note="27 copies 2 mer to 85% conserved"				
repeat_region	767..810				
	/note="11 copies 4 mer tctc 93% conserved"				
repeat_region	813..1075				
	/note="AluJB repeat: matches 1..274 of consensus"				
repeat_region	1166..1204				
	/note="THE1B repeat: matches 326..364 of consensus"				
repeat_region	1205..1504				
	/note="AluSg repeat: matches 1..301 of consensus"				
repeat_region	1505..1830				
	/note="THE1B repeat: matches 1..326 of consensus"				
misc_feature	1974..2325				
	/note="match: 267843 STS containing (CA) repeat"				
repeat_region	2115..2143				
	/note="14 copies 2 mer ca 100% conserved; differs from 267843"				

repeat_region	28124..28262	/rpt_family="MER21_g"
repeat_region	28335..28657	/rpt_family="MER21_g"
repeat_region	28927..28969	/rpt_family="L2"
repeat_region	30119..30461	/rpt_family="MaLR"
repeat_region	30476..30604	/rpt_family="MIR"
repeat_region	30989..31141	/rpt_family="MIR"
misc_feature	complement(31189..31236)	/note="match to EST (NID:92073201)"
repeat_region	31207..31482	/rpt_family="Alu"
repeat_region	31627..31721	/rpt_family="MIR"
repeat_region	31855..31927	/rpt_family="MIR"
repeat_region	32385..32584	/rpt_family="L2"
repeat_region	32740..33139	/rpt_family="Retroviral"
repeat_region	33244..33314	/rpt_family="MIR"
repeat_region	33315..33621	/rpt_family="Alu"
repeat_region	33622..33745	/rpt_family="MIR"
repeat_region	34116..34418	/rpt_family="Alu"
repeat_region	36139..36443	/rpt_family="Alu"
repeat_region	36572..37062	/rpt_family="L2"
repeat_region	37202..37316	/rpt_family="MER1_type?"
repeat_region	37836..37899	/rpt_family="MER1_type?"
repeat_region	38128..38372	/rpt_family="L2"
repeat_region	39143..39335	/rpt_family="MIR"
repeat_region	39481..39628	/rpt_family="MIR"
repeat_region	39846..39890	/rpt_family="L2"
repeat_region	39917..40378	/rpt_family="L1"
repeat_region	40379..41802	/rpt_family="L1"
repeat_region	41853..41939	/rpt_family="L1"
repeat_region	41949..42017	/rpt_family="L2"
Query Match	14.1%; Score 33; DB 11; Length 116218;	
Best Local Similarity	51.0%; Pred.No. 8;	
Matches	78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;	
QY	74	CTTTGCAACTGAAGTGAAGGAGTCTTTCTGACAAATTCCTCTATGAGTCCAGCTTCC 133
DB	81977	CTCTGTGTATAGATTAAGGCTTAATCTCTGACACTTTGGTCTTCCTCTAAACCAT 81918
QY	134	TGGAATTCCTGGAAGCTCTGCCTCTCCCTCCATTCCTCAGGACCGCTCAACC 193
DB	81917	CCTACTGCTTTCAAGAGTTCAAGTCTCTCTCTGTCCTCCATTTCCAGTCTCAGTTTCA 81858
QY	194	TCCACCATGCAAGATCTCAACACCATGTTGTCT 226
DB	81857	TCGTAAATGCAGGACTTGACITTTATGTTTCT 81825

RESULT	10	
AB014077/c		
LOCUS	AB014077 30452 bp DNA PRI 20-NOV-1999	
DEFINITION	Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region, Cosmid clone:TY7H8, complete sequence.	
ACCESSION	AB014077	
VERSION	AB014077.1 GI:5672586	
KEYWORDS	HTG.	
SOURCE	Homo sapiens (haplotype:HLA-DR4) adult male blood Immunoresponse cell cell_line:Boleth DNA, clone_lib:CEPH YAC clone:745D12 and 960H11.	
ORGANISM	Homo sapiens	
REFERENCE	1 (sites)	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T., Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y., Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A., Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M., Bahram,S. and Inoko,H.	
TITLE	Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)	
MEDLINE	20027539	
REFERENCE	2 (bases 1 to 30452)	
AUTHORS	Shina,T., Tamiya,G., Oka,A. and Takishima,N.	
TITLE	Submitted (18-MAY-1998) to the DBJ/EMBL/GenBank databases.	
JOURNAL	Nobusada Takishima, Tokai University School of Medicine, Molecular Life Science; Bohseidai, Isehara, Kanagawa 259-11, Japan [E-mail:takisand@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121, Fax:81-463-94-8884]	
FEATURES	Location/Qualifiers	
	1..30452	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/cell_line="Boleth"	
	/cell_type="Immunoresponse cell"	
	/chromosome="6"	
	/clone="745D12 and 960H11"	
	/clone_lib="CEPH YAC"	
	/dev_stage="adult"	
	/haplotype="HLA-DR4"	
	/map="6p21.3"	
	/sex="male"	
	/tissue_type="blood"	
	/note="Cosmid clone TY7H8"	
BASE COUNT	7732 a 6763 c 7522 g 8435 t	
ORIGIN		
Query Match	14.0%; Score 32.8; DB 10; Length 30452;	
Best Local Similarity	58.0%; Pred. No. 8.5;	
Matches	58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;	
QY	64	CGGGTATGATCTTTCGAACCTGAAGGAGTCTTTCTGACAAATTCCTCTATGAG 123
DB	12221	CGGCCCTGCTCCACCACCTCCAGCTCCAGCATCTACTTAAGCAACTCTTATTTCT 12162
QY	124	TCCAGCTTCTCGAATTCCTTGAAGAGCTCTGCCCTCC 163
DB	12161	CCTAGATTACTGAATAGCTCAACTGCTCTCTCTGCTCC 12122

RESULT	11
AP000513/c	
LOCUS	AP000513 100000 bp DNA PRI 28-SEP-1999
DEFINITION	Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region, section 12/20.
ACCESSION	AP000513
VERSION	AP000513.1 GI:5926700
KEYWORDS	

SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (sites)
JOURNAL Shinya S., Tamiya, G., Oka, A. and Inoko, H.
REFERENCE Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
AUTHORS published Only in Database (1999) in press
TITLE 2 (bases 1 to 100000)
JOURNAL Hirakawa, H., Imaquuchi, H., Imai, K. and Shimada, J.
REFERENCE Direct Submission
AUTHORS Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
TITLE Hirakawa, Japan Science and Technology Corporation (JST), Advanced
JOURNAL Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
REFERENCE Japan [E-mail: mika@tokyo.jst.go.jp].
AUTHORS URL: http://www.alis.tokyo.jst.go.jp/. Tel: 81-3-5214-8491,
TITLE Fax: 81-3-5214-8470)
JOURNAL This sequence is conducted by Tokai University as a JST sequencing
COMMENT Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone: +81-463-93-1121, Fax: +81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive web site (<http://www.alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www.alis.tokyo.jst.go.jp.
FEATURES
Source
1..100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
/complement(25387..25469)
/note="316A/316S: The location is between each flanking
STS site of PCR primers."
/db_xref="GBA:215649"
/complement(36020..36262)
/note="TIGR-A007C28: The location is between each flanking
STS site of PCR primers."
/db_xref="GBA:4588360"
/complement(36173..37150)
/note="SGC34051: The location is between each flanking site
of PCR primers."
/db_xref="GBA:4575750"
/complement(41454..41670)
/note="stSG3452: The location is between each flanking site
of PCR primers."
/db_xref="GBA:4573047"
56458..56539
/note="960R: The location is between each flanking site of
PCR primers."
/db_xref="GBA:1297691"
56453..56581
/standard_name="D6S2266"
/note="sy960h11-B: The location is between each flanking
STS site of PCR primers."
/db_xref="GBA:5583962"
80448..80597
/note="SGC-32967: The location is between each flanking
STS site of PCR primers."
/db_xref="GBA:6456197"
/complement(96072..96222)
/note="W1-17348: The location is between each flanking site
of PCR primers."
/db_xref="GBA:4580894"
/complement(96558..96659)
/note="2185/2186: The location is between each flanking
STS site of PCR primers."
/db_xref="GBA:581290"
25730 a 23716 c 23917 g 26637 t
"GIN"

Query Match 14.0%; Score 32.8; DB 10; Length 100000;
Best Local Similarity 58.0%; Pred. No. 9.1;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 64 CGGGTATGACTTCCCACTGAAGCTGAAGAGTCTTTCTGACAAAATTCCTCTATGAG 123
Db 91709 CGGGCTCTGTCACCACTCCAGCTCCAGCTCACTACTAAGCAACTCTTATTTCT 91650
QY 124 TCCAGCTTCTGGAATGCTTGAAGAGCTCTGCTCTCTCC 163
Db 91649 CCTAGATTACTGAATAGCCTCACTGCTCTCTCTCTCC 91610
RESULT 12
AC004209/c
LOCUS Homo sapiens clone UWGC:y55c025 from 6p21, complete sequence. PRI
DEFINITION AC004209
ACCESSION AC004209
VERSION AC004209.1 GI:3522963
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 43600)
Janer, M., Guillaudoux, T., Vu, Q., Kutyavin, T., Harter, H. and
Geraghty, D.E.
TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
2 (bases 1 to 43600)
Geraghty, D.E. and Olson, M.V.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1998) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 43600)
AUTHORS Geraghty, D.E. and Olson, M.V.
JOURNAL Direct Submission
Submitted (03-SEP-1998) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, Seattle, WA 98195, USA
REMARK The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
Contact: Daniel E. Geraghty (geraghty@fhcr.org)
On Sep 3, 1998 this sequence version replaced gi:2905851.
COMMENT Overlapping Sequences:
5': UWGC:y55c174
3': UWGC:y55c068

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
Double stranded (DS) coverage: 79.5%
DS or two chemistry coverage: 99.7%
Single stranded regions: 2

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
Mapping. Comparison of the experimentally derived map digest
fragments with sequence-predicted fragments is given below.

Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

Map	Seq	HindIII	Map	Seq
5990.87	5842.00	1561.21	1547.00	2899.98 2857.00
1246.59	1240.00	1365.20	1360.00	4696.59 4486.00
628.61	627.00	1803.37	1799.00	8400.84 8055.00
12412.85	11969.00	13189.91	12685.00	10629.14 10329.00
9189.25	8932.00	4298.12	4138.00	951.62 941.00
2647.86	2596.00	736.47	727.00	7859.84 7632.00
989.83	992.00	2883.77	2832.00	
988.83	977.00	1210.44	1211.00	
6670.61	6509.00	8305.16	8204.00	
		7741.77	7605.00	

FEATURES

Source	Location/Qualifiers
1	43600
1	organelle="Homo sapiens"
1	dbxref="tax:9606"
1	chromosome="8"
1	clone="CGM:DL25C12"
1	sub_clone="UMGC:Y55C025"
1	cell_line="CGM1"
1	clone_lib="Wash U YAC Library"
1	map="6p21"
repeat_region	complement(21..277)
repeat_region	/rpt_family="Alu"
repeat_region	676..873
repeat_region	/rpt_family="Alu"
repeat_region	complement(1150..1448)
repeat_region	/rpt_family="Alu"
repeat_region	2056..2345
repeat_region	/rpt_family="Alu"
repeat_region	complement(2630..2913)
repeat_region	/rpt_family="Alu"
repeat_region	complement(6636..6708)
repeat_region	/rpt_family="Alu"
repeat_region	complement(6854..7129)
repeat_region	/rpt_family="Alu"
repeat_region	complement(7265..7543)
repeat_region	/rpt_family="Alu"
repeat_region	complement(7635..7770)
repeat_region	/rpt_family="Alu"
repeat_region	complement(8253..8535)
repeat_region	/rpt_family="Alu"
repeat_region	complement(8973..9090)
repeat_region	/rpt_family="Alu"
repeat_region	15507..15767
repeat_region	/rpt_family="Alu"
repeat_region	17528..17816
repeat_region	/rpt_family="Alu"
repeat_region	18354..18646
repeat_region	/rpt_family="Alu"
repeat_region	complement(19460..19743)
repeat_region	/rpt_family="Alu"
repeat_region	complement(26422..26590)
repeat_region	/rpt_family="MER42"
repeat_region	complement(26722..27017)
repeat_region	/rpt_family="Alu"
repeat_region	complement(27496..27795)

repeat_region	/rpt_family="Alu"	complement(27976..28261)
repeat_region	/rpt_family="Alu"	complement(28394..28664)
repeat_region	28716..28998	
repeat_region	/rpt_family="Alu"	complement(29874..29952)
repeat_region	/rpt_family="Alu"	complement(30110..30353)
repeat_region	complement(30451..30660)	
repeat_region	/rpt_family="Alu"	complement(30685..30865)
repeat_region	/rpt_family="MER42"	complement(30862..31150)
repeat_region	/rpt_family="Alu"	complement(31182..31468)
repeat_region	/rpt_family="Alu"	complement(31482..31678)
repeat_region	/rpt_family="MER42"	31795..32069
repeat_region	/rpt_family="Alu"	32112..32243
repeat_region	/rpt_family="Alu"	complement(32820..33022)
repeat_region	/rpt_family="MLT2B2"	33228..33415
repeat_region	/rpt_family="MLT2B2"	33680..33971
repeat_region	/rpt_family="Alu"	34331..34939
repeat_region	/rpt_family="Alu"	35200..35499
repeat_region	/rpt_family="Alu"	35906..36190
repeat_region	/rpt_family="Alu"	complement(36377..36494)
repeat_region	/rpt_family="LTR8"	36578..36830
repeat_region	/rpt_family="Alu"	37011..37330
repeat_region	/rpt_family="Alu"	complement(37530..37676)
repeat_region	/rpt_family="LTR8"	complement(37982..38276)
repeat_region	/rpt_family="Alu"	complement(38785..39164)
repeat_region	/rpt_family="Alu"	complement(39725..40004)
repeat_region	/rpt_family="Alu"	complement(40229..40521)
repeat_region	/rpt_family="Alu"	complement(40667..43600)
repeat_region	/rpt_family="L1"	

BASE COUNT 11071 a 9661 c 10338 g 12530 t
ORIGIN

Query Match 14.0%; Score 32.8; DB 11; Length 43600;

Best Local Similarity 58.0%; Pred.No.8.7; Indels 0; Gaps 0;

Matches 58; Conservative 0; Mismatches 42;

Oy 64 CCGGGTATGACTTGCAGCTGAAGAGTCTTTTCTGACAAATCTCTCTATGAG 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 17165 CCGGCCCTGTCTCAGCAGCTCCAGCATCTACTAGCAACTCTTTCTCT 17106
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 124 TCCAGCTTCTCGGAATGTCTTGAAGAGCTCTGCTCTCTCC 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 17105 CTTAGATTAAGTAAAGCTCAACTGCTCTCTCTGCTCC 17066
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13

```

AC002327      AC002327      15915 bp      DNA
DEFINITION    Mus musculus chromosome 7, clone 19K5, complete sequence.
ACCESSION     AC002327
VERSION       AC002327.1 GI:4210518
KEYWORDS      HTG.
SOURCE        house mouse.
ORGANISM      Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS       Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
TITLE         1 (bases 1 to 15915)
JOURNAL       Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE     Mus musculus chromosome 7, clone 19K5
AUTHORS       2 (bases 1 to 15915)
Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,
Fasman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,
Barnan, N., Chang, A., Cooke, P., Daly, M.J., Devon, K., Dewar, K.,
Forrest, C., Gage, D., Geraigery, K., Guitau, G., Hagos, B., Huang, J.,
Jacotot, L., Lane, M., Lee, K., MacKenzie, J., Marquis, N.,
McBroom, J., Molla, M., Moloney, N., Morrow, J., Nachman, A.,
Naylor, J., O'Connor, T., Olotu, A., Peterson, K., Rollins, G.,
Spencer, J., Stilwell, J., Stone, C., Strickland, C., Sydney, K.,
Traish, A., Wilmer, F., Zentseva, I. and Zody, M.
Direct Submission
Submitted (15-JUL-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE     3 (bases 1 to 15915)
AUTHORS       Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Beun, J., Brown, A.,
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K.,
Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
Funk, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hagos, B., Heaford, A., Horton, L., Howland, J.C., Jones, C., Kann, L.,
Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
Marquis, N., McEwan, P., McKernan, K., McKernan, K., McLaughlin, J.,
Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Nijoff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Tasfaye, S., Toruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W.J. and Zody, M.
Direct Submission
Submitted (02-FEB-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT       On Feb 2, 1999 this sequence version replaced gi:4204708.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html.
FEATURES     source
              Location/Qualifiers
                1..15915
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="19K5"
                /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
                1-194)"
                /map="7"
                /chromosome="7"
                complement(81..193)
                /rpt_family="Alu/FLAM"
                718..860
                /rpt_family="(TA)n"
                863..942
                /rpt_family="AT_rich"
                1099..1151
                /rpt_family="AT_rich"
                1482..1542
                /rpt_family="FRAM/FLAM"
                1753..1782
                /rpt_family="AT_rich"
                1908..1957
                /rpt_family="Alu"
                2052..2128
                /rpt_family="AluSg/x"
                2709..2759
                /rpt_family="AluSg"
                2806..2832
                /rpt_family="(A)n"
                2852..2914
                /rpt_family="MIR"
                2996..3022
                /rpt_family="AT_rich"
                3827..3913
                /rpt_family="L2"
                4603..4679
                /rpt_family="(CGG)n"
                4722..4817
                /rpt_family="(CGG)n"
                4832..4954
                /rpt_family="GC_rich"
                6796..6818
                /rpt_family="(TCCC)n"
                6964..7138
                /rpt_family="MERSA"
                7154..7191
                /rpt_family="(TC)n"
                complement(8423..8550)
                /rpt_family="MIR"
                8635..8751
                /rpt_family="MIR"
                complement(9506..9667)
                /rpt_family="AluSg/x"
                complement(9699..9812)
                /rpt_family="L2"
                10967..11107
                /rpt_family="FRAM"
                11756..11822
                /rpt_family="(TTTTA)n"
                complement(11874..11966)
                /rpt_family="AluSg"
                12017..12047
                /rpt_family="(TTTA)n"
                complement(12138..12196)
                /rpt_family="tRNA-Lys-AAA"
                12402..12515
                /rpt_family="MIR"
                12665..12749
                /rpt_family="FLAM_A"
                complement(12899..12968)
                /rpt_family="MIR"
                13967..14023
                /rpt_family="(TCCC)n"
                14765..14811
                /rpt_family="AT_rich"
                complement(15408..15604)
                /rpt_family="MIR"
                complement(16028..16174)
                /rpt_family="FLAM_C"
                18347..18468
                /rpt_family="MIR"
                complement(18522..18577)
                /rpt_family="MIR"
                complement(19193..19405)
                /rpt_family="AluJb"
                19549..19601
                /rpt_family="PWERI"
                23117..23147
                /rpt_family="AT_rich"
                complement(23988..24046)
                /rpt_family="tRNA-Ala-GCG"
                24457..24544
                /rpt_family="FLAM_A"
                25145..25204
                /rpt_family="MIR"
                25289..25322

```

```

repeat_region      /rpt_family="AT-rich"
                   complement(25541..25624)
repeat_region      /rpt_family="L2"
                   25694..25743
repeat_region      /rpt_family="(TTCA)n"
                   26325..26466
repeat_region      /rpt_family="AluSg"
                   27277..27299
repeat_region      /rpt_family="(TTTG)n"
                   complement(27302..27443)
repeat_region      /rpt_family="FLAM_C"
                   complement(27608..27652)
repeat_region      /rpt_family="MIR"
                   complement(28409..28532)
repeat_region      /rpt_family="MIR"
                   complement(30322..30465)
repeat_region      /rpt_family="L2"
                   35588..35800
repeat_region      /rpt_family="AluJ"
                   40654..40703
repeat_region      /rpt_family="A-rich"
                   42215..42257
repeat_region      /rpt_family="CT-rich"
                   43511..43544
repeat_region      /rpt_family="AT-rich"
                   43715..43742
repeat_region      /rpt_family="(T)n"
                   43745..43767
repeat_region      /rpt_family="AT-rich"
                   43817..43856
repeat_region      /rpt_family="AluJ"
                   44196..44221
repeat_region      /rpt_family="(CCCTAA)n"
                   44395..44471
repeat_region      /rpt_family="AluSp"
                   46141..46271
repeat_region      /rpt_family="FRAM"
                   46603..46641
repeat_region      /rpt_family="AT-rich"
                   47045..47073
repeat_region      /rpt_family="(CA)n"
                   47293..47357
repeat_region      /rpt_family="(GA)n"
                   complement(48842..49005)
repeat_region      /rpt_family="FRAM"
                   49023..49127
repeat_region      /rpt_family="AluSp"
                   49498..49629
repeat_region      /rpt_family="FLAM_C"
                   complement(50407..50479)
repeat_region      /rpt_family="L2"
                   50860..50920
repeat_region      /rpt_family="AluS"
                   50929..50965
repeat_region      /rpt_family="(GTTTG)n"

Query Match      14.0% Score 32.8; DB 12; Length 159515;
Best Local Similarity 53.0%; Pred. No. 9.4;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 99 TTTTCGACAAATCCCTCCTATGATGCTTCCTGGAATGCTTGAAAGACTCTCGCT 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52401 TTTTACTGTTTATTTTAAAAAGAGCCCAACTTGTCTATGATGACCAAGGATGACTGGACTT 52460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 CCTCCCTCCATCCTCTTCAGGAGCAGCGTCACCTCCACCATGCAAGATCTCAACACCA 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52461 CCTGTCCACCTCTCAGTTGGAGAGTCGGTCAGGTACGCCATCCCTATATCTCTACGCT 52520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 TGTGTCTGCAC 230
   ||| ||| |||
Db 52521 GGCTCTTACCA 52532
   ||| ||| |||

```

RESULT 14 HS179N16/c

LOCUS
DEFINITION

HS179N16 172048 bp DNA PRI 23-NOV-1999
Homo sapiens DNA sequence from PAC 179N16 on chromosome
6p21.1-21.33. Contains the SAPK4 (MAPK p38delta) gene, and the
alternatively spliced SAPK2 gene coding for CSaids binding protein
CSBP2 and a MAPK p38beta LIKE protein. Contains ESTs, STSs and two
predicted CpG islands, complete sequence.

ACCESSION
VERSION
KEYWORDS

295152 1 GI:3036773
HTG, CpG island; CSaids binding protein; CSBP2; MAPK; Mitogen
activated protein kinase; MXI2; P38; SAPK4; Stress activated
protein kinase.

SOURCE

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172048)

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (02-APR-1998) Chromosome 6 Project Group
(http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquig@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Apr 8, 1998 this sequence version replaced gi:2505919.
IMPORTANT: This sequence is the entire insert of clone 179N16.
During sequence assembly data are compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed in collaboration by the Sanger
Centre chromosome 6 mapping group and Ioannis Ragnousis. Further
information can be found at http://www.sanger.ac.uk/HGP/Chr6/ This
sequence has been finished according to sequence map criteria as
follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

The true left end of clone 179N16 is at 1 in this sequence. The
true right end of clone 524E15 is at 630.
The true left end of clone 179N16 is at 172048.
179N16 is from the library RCI1 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.

For further details see http://bacpac.med.buffalo.edu/
Location/Qualifiers

FEATURES

source
1..172048
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p21.1-21.33"
/clone="RPI-179N16"
/clone_lib="RCI-1"
complement(1..1077)
/notes-match: STS AL040972"
4..457
/note="11 repeat: matches 5082..4624 of consensus"
493..734
/note="AluJo repeat: matches 302..62 of consensus;
incomplete repeat"
1714..1954
/note="L1MB8 repeat: matches 324..574 of consensus"
1956..2257
/note="AluJb repeat: matches 1..302 of consensus"
2258..2543
/note="L1MB6 repeat: matches 566..874 of consensus"
2548..2848
/note="AluSx repeat: matches 301..1 of consensus"
2879..3160
/note="AluJo repeat: matches 277..2 of consensus";

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```

repeat_region      incomplete repeat"
3168..3369
/note="MER44C repeat: matches 258..463 of consensus"
4208..4497
/note="AluJb repeat: matches 289..1 of consensus"
5473..5894
/note="MLT1C repeat: matches 39..465 of consensus"
5986..6284
/note="AluJb repeat: matches 1..296 of consensus"
6361..6420
/note="AluSx repeat: matches 301..242 of consensus;
incomplete repeat"
6612..6914
/note="AluJb repeat: matches 2..302 of consensus"
6951..7084
/note="FLAN_C repeat: matches 1..133 of consensus"
7621..7941
/note="L1 repeat: matches 5068..5387 of consensus"
7809..8170
/note="L1MC2 repeat: matches 15..384 of consensus"
8255..8378
/note="MER2 repeat: matches 13..136 of consensus"
8396..8603
/note="AluJb repeat: matches 291..82 of consensus;
incomplete repeat"
8605..8691
/note="MER2 repeat: matches 125..209 of consensus"
8826..8890
/note="MER2 repeat: matches 279..344 of consensus"
9518..9894
/note="THE1C repeat: matches 3..371 of consensus"
9964..9989
/note="13 copies of 2 mer 96 % conserved"
10970..11269
/note="AluS9 repeat: matches 300..1 of consensus"
12207..12333
/note="MIR repeat: matches 106..248 of consensus"
12735..13031
/note="AluSx repeat: matches 1..298 of consensus"
13042..13256
/note="AluSx repeat: matches 2..217 of consensus;
incomplete repeat"
13494..13581
/note="AluSx repeat: matches 205..288 of consensus;
incomplete repeat"
13595..13953
/note="THE1B repeat: matches 364..1 of consensus"
13972..14648
/note="L1MC3 repeat: matches 1676..2345 of consensus"
14649..14948
/note="AluSx repeat: matches 1..301 of consensus"
14949..15077
/note="L1MC3 repeat: matches 2336..2470 of consensus"
15102..15231
/note="FLAN_A repeat: matches 133..4 of consensus"
15244..15289
/note="23 copies of 2 mer 100 % conserved"
15674..15853
/note="match: Z63849 Cpg Island clone 91c11"
16341..16702
/note="MLT2FA repeat: matches 33..394 of consensus"
16823..16969
/note="MER11A repeat: matches 4..151 of consensus"
16968..17594
/note="MER11A repeat: matches 60..738 of consensus"
17278..17886
/note="MER11B repeat: matches 2..627 of consensus"
17977..18027
/note="MLT2FA repeat: matches 400..450 of consensus"
19217..19333
/note="MIR repeat: matches 262..117 of consensus"
19471..19693
/note="AluJo repeat: matches 85..301 of consensus;

```

```

repeat_region      incomplete repeat"
21094..21432
/note="MER1B repeat: matches 1..337 of consensus"
21701..22077
/note="match: ESTs R71395 R69756 AA587582 AA629092 R50348"
21706..21985
/note="match: STS G23312"
22054..22225
/note="AluJb repeat: matches 131..301 of consensus;
incomplete repeat"
22217..22628
/note="match: ESTs R50012 R70180"
24456..24530
/note="MIR repeat: matches 41..123 of consensus"
24535..24834
/note="AluSx repeat: matches 302..1 of consensus"
24840..25118
/note="L1MC2 repeat: matches 1065..772 of consensus"
25136..25452
/note="AluJb repeat: matches 11..302 of consensus"
25649..25765
/note="MER5A repeat: matches 63..189 of consensus"
26130..26545
/note="MLT1B repeat: matches 1..390 of consensus"
complement(<26546..35701)
/gene="SAPK4"
/note="match: I63675 patent US 5663313"
complement(join(26546..26932,27171..27347,27446..27524,
27782..27861,29276..29347,29456..29570,29695..29742,
30178..30207,30365..30473,33547..33605,34826..34955,
35525..35905))
/gene="SAPK4"
/note="match: cDNAs U93232 U66243 Y10487 AF015256 Y10488
AF004709"
/evidence=not_experimental
/product="dj179N16.1 (Stress Activated Protein Kinase 4
(Mitogen Activated Protein Kinase p38delta))"
/gene="SAPK4"
complement(26546..35905)
complement(join(26853..26932,27171..27347,27446..27524,
27782..27861,29276..29347,29456..29570,29695..29742,
30178..30207,30365..30473,33547..33605,34826..34955,
35525..35643))
/gene="SAPK4"
/note="match: proteins O14739 O15124"
/codon_start=1
/evidence=not_experimental
/product="dj179N16.1 (Stress Activated Protein Kinase 4
(Mitogen Activated Protein Kinase p38delta))"
/protein_id="CA808438.1"
/db_xref="GI:3036774"
/db_xref="SPTREMBL:O15264"
/translaton="MSLIRKGFYQDVNKTAVELPKTVSPTHVSGAGSYVCSAID
KRGSKVAIKLSRFQSEIFAKRAYRELLLLKMHQENNVIGLDVETPASSIRNFYD
FYLVPFMOTDLOKIMGMEFEKIOVLVOMLKGLKYHSAGVYVHRDLAPGRLAYNE
DCELKILDFGLARHADAEMTGYVTRYRAPEVILSWHYNGTVDINSVCINAEMLT
Query Match      13.9%; Score 32.6; DB 10; Length 172048;
Best Local Similarity 55.9%; Pred. No. 11;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 73 ACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCTAGTCCAGCTTC 132
Db 12233 ACAAGAAGACTGAATCTCTAAGAGTTTTTTTTTCTTANTCCCAAGTTTTCTCTTA 12174
QY 133 CTGGAATTGCTGAAGACTCTGCCTCCTCCATCTCCCTTCAGGACC 183
Db 12173 CTGCAAGTTTTCTTTTCAGTCTCTCCACCAAGCAACTCTCTGATGCATC 12123

```

```

RESULT 15
EHYU20824/c
LOCUS      EHYU20824 184427 bp DNA VRL 02-FEB-1996.

```

Wed Mar 22 12:07:50 2000

```

DEFINITION   Equine herpesvirus 2, complete genome.
ACCESSION    U20824
VERSION      U20824.1  GI:695172
KEYWORDS
SOURCE       Equine herpesvirus 2.
ORGANISM     Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
              Bacteriophirinae;
REFERENCE    1 (bases 1 to 184427)
              Telford,E.A., Watson,M.S., Aird,H.C., Perry,J. and Davison,A.J.
              The DNA sequence of equine herpesvirus 2
              J. Mol. Biol. 249 (3), 520-528 (1995)
              95302501
REFERENCE    2 (bases 1 to 184427)
              Telford,E.A.R.
              Direct Submission
              Submitted (09-FEB-1995) Elizabeth A.R. Telford, Institute of
              Virology, MRC Virology Unit, University of Glasgow, Church Street,
              Glasgow G11 5JR, United Kingdom
              Location/Qualifiers
              1..184427
              /organism="Equine herpesvirus 2"
              /strain="86/67"
              /db_xref="taxon:12657"
              1..17553
              /rpt_type=direct
              9248..10399
              /note="ORF E1"
              /codon_start=1
              /product="g protein-coupled receptor"
              /protein_id="AAC13788.1"
              /db_xref="GI:695173"
              /translation="MATTSAITSTVNTSSLTATMTNTFTSLTSTVVTIASLVPSTNNS
              EBYDLDLDVDYDEAPCYKSDTRLAQVVALYLLVFLFGLIGNILVIVIRYKH
              LKNTMLLMALISDLFLTLFPMWYHGMTHDTEGSLKRLRGVCYMSLYSQV
              FCILLVDRYLAVVYATLRFVTCVCTWELAGLISLPEFFFHQDDNG
              RVQCDPIYPMKSNVWRARVAVIMLSLILPILMAVCYVIVIRLLRRPKKYKA
              IKLFIYAVYVFWPTPNIVLLSTETATLNLQALSSNLDMLLITKTAYTHCC
              INPIYAFGEKERHLLHFFHTYVAILGCKYIPLSGDGEGKEGPRI"
              10494..10499
              /note="ORF E1"
              10660..10665
              /note="ORF E1"
              12533..13114
              14197..14541
              18810..19733
              /note="ORF E2, product has hydrophobic N-terminus"
              /codon_start=1
              /protein_id="AAC13789.1"
              /db_xref="GI:695174"
              /translation="MIILGLVLSLASTASATQAMPQVITGHEGRPLSVRCHAGK
              NFKPTMLLTKTIEGSAFRGTVIPFRKGPVYTKSSSGGTGFTVDSLMDAGRYA
              CAVSQENDPKDESQSYHTATNTIVLRRRRSLGLPQTPAGRSFQLSQTGP
              LSSQSESVVAVKGVKVVGSSLHVECOFSGATAGLNISAVLFKVGEGYQAVST
              ISTSLITLDSLVPIETGQAKCGVNATKEDSGEYMCNIGMDPPEGESEGDGDFSG
              VDFEKLITVTERGDEADDDLYLTSSEKK"
              19761..19766
              /note="ORF E2"
              19784..19789
              /note="ORF E2"
              20320..20844
              /note="ORF E3, product has N- and C- hydrophobic domains
              and 6 N-linked glycosylation sites"
              /codon_start=1
              /protein_id="AAC13790.1"
              /db_xref="GI:695175"
              /translation="MATEQFACALVWLIGCYAGTPTSKSTLIYNSTQVPSIE
              NGQSLIYNGSWFIYEFNCSGIEQLQWPTTICFPANPANIWNGPFCYKLNIT
              TLMPTSTSTVTTGIFPDQNTHTHTYKTRPNLLRFGITPWAITLIVIL
              LVVIYVNCMPMF"
              21360..25319
              /note="ORF 03"
              /codon_start=1

```

```

/product="tegument protein"
/protein_id="AAC13791.1"
/db_xref="GI:695176"
/translation="NDSTSKRSFKRAGGRATVYFVPTWQHGGQGPVSSFAFSL
VNEVGFGLDYGCAGCTGATAGRAAAVDPGVASEAFMEILRTVLPGHVGAVSSELP
AGLSYLTSGDAGAGPCTLSDDVPMWOLLEHLIRGOOOLRRPLVGDAGEAGRGP
GEGEATPIPKILAYSEFFHTLSLSIKOHAYMLGNOLSSPASCFTPGYLGFTPLG
DGERPDLKFARVPFALAHKADAKLPEFNTREHSDVLTGTHGCAIKRPNGL
LGYCLRYAPTSRVFIDRVFSSNGKPDIFGVYQVGGISGASELDLLISSEAGFGW
LGLVGLHGVVGGSEDFALQYLESHLAQRCGAPTLGLFVRVMDSGSPFGMGPA
SRPHSQTGATRASTIISPKACLSPPVRGGGASVFMGLDFQFQRRAGGAYASHDGRSLA
SILAAYQFTINAGGRAVLTSCQRLGTHAKIAAACGVMLGKTYTSLPEAVSEL
SHEPQNASMNRSTILLSKFLNVEAQVLLTVNTPNRQILEGVTYKTFGRVSLGHL
ABAKIYLDSDSPDYCOGPRYLCFKMKHPRYALPIDRMCTASTYTPASRVPYNS
SDAGPPSDAEVLLGILRPTVGCNKYIVKHVDRCAAGVACQCGVGLDLPVADYSM
VYVESAGVPSSEPTWETROVDKCELIHEDGVSGVCSALGKGVFPFYPEAGAK
MALRESUSLALAPARIETEDVYNLSLAWPHILGAQGEVFTLLPCKRFCEADGVCN
VTSASSRAGGDDPPEGADIKSFVASAFATPTPCAFKWLTPDLKEARSVLEPLASP
GHAFASYOQVGRKKYTGSPVNISSGAYLAKLARAVLHLRAKRLVSGHDVGDGLW
AALAEMANMRSIDVEVPVRLDGMGLTSETPGVALEVPSSASAAYEILRSKIY
LCYRVGTGHTKTRFRITRGDQVLECEQEVEGEAMRSHSRMERTGCFNRHAYG
EGPYERERAEFLAYTPFRLNPKKHRAVALLPGCPTPLAALDRLSGFDALAYS
HLDLAHVLTVDEVCGLFVAGVDNTSDTKTSDALTSALVHSRNEGVIRDFLSKDTFSLG
VSMACRILFDGMAYSPALQSRDRCVPVASCCKFESRMLNFIQSTPAVAFETLRG
SILPCWAGTHLGFGRASPECIDRLVASGOAASMYGSDVASGPAMAYPANPTETLPV
AGICSDGRHLVLDLHIDITASYILWQWPHVPKSNIPINAVSPMKQVQVYDLHANVTAHSH"
25639..26190
/note="ORF E4"
/codon_start=1
/protein_id="AAC13792.1"
/db_xref="GI:695177"
/translation="MVPSEFFFEEDRVRLENEAQKLSLTNLLKNVFAQTLDKMKPEGV
LTFTEALLAWVDECKEYELHOLIELVMQVPSVEAFTSAINSLGIRIOTHQGGED
NFRRLCSISFASCFELVLDQSDERCLSVFASLAKAYVESQNLMLAYSGLSAGLNE
RFRSMWLFALKAQKWLRFITFFK"
26436..269873
/note="ORF 06"
/codon_start=1
/product="single-stranded DNA binding protein"
/protein_id="AAC13793.1"
/db_xref="GI:695178"
/translation="MNSNRAAGPVEENVGTQASVGCPCGVYLYPGDTFFPVEEASLIG
NLHAGVEFSLPLLSGLVADPVENNVKAVHKLDPATVSKASAYHREVTVFANAAC
FKETPAGLEGLECAASROLGVAPEERAGAGAPPELADLGLLPGAEGSHIAGVTV
TESFKERLYGQLYVYESQVTSVYGECDKFKVPLVDGELEFASPCRENLRVYFVHAGV
SYLFEAHTLSLAQRVDRVPGILGALERSFHDQYKLPKYVECPREPATGHRGAD
CSLTIVDSVATELAVYGLSEFLEVPQEGTALLSDKMPFEGCETPEQVREALQFNA
KQAVHYHSQLSFSGNSVLYLARVQKQNSRGGENVYNSFFMGHGLACLAREPTQKNG
LPSPGPVAPASLGSNTSLHLAYAAASFQNLARHCYILQFAHQKRSNNSGINFT
YVGTAAITPMLCCLRGSCPCVNTLFLYRLDRFPVAVSVRRDPTVTVVAGAYNDL
DIAGNANYRQDEESNQBSEKFTYQVOTVTLERLSEAGICEGEGDGAHNTG
SFLYKPEIDGLVDEVARFTNSMKNVNYRESIKSIHHVQVVCNYYQVPCPFVFL
NLYRCVLAVQDLCITCMYQENPANGVSPGEMLMKHVQTLWTNFKNSCIDKGLV
TGTEIKRVHNDQCFDFDSDAARGFVSCKTOVLSALMAPRVNKKNIIFSN
FYFVFNPRVSELANVYEQNSRLHGETSIIDVWFENFYIYAKVRLNNALEFRACQTO
FYATTIHLTPKIQTPAEYHALGARGVADVGEIIGAAKELVTVTVGTSRDNICE
VSKCRPIVLPVNVKTYVTGNSQIFQCANGLYFGRGVKKNLIPDGSFKKQGVST
WKKRRHVFTPLVNLKRSYQGAFAFEIEGVRVRRVQOILSDGDNPHVIRVQLQV
LSIGSECRSVSEVDLEYTMQYIYFAGDYSERLQSLDGLGDWSEWALSVLGEEDP
LGLSELEKVEDAECLGHPQDDEPALAPQAAAPQYSGSSSVAGKRRKANVILGDLDL"
29997..32183
/note="ORF 07"
/codon_start=1
/function="role in capsid maturation/DNA packaging"
/protein_id="AAC13794.1"
/db_xref="GI:695179"
/translation="NARELAAYAAQVFDLAAEYSLIGYCDFSSIDKRCVMANSKNVFK
LCSELCLLRQNDTECSPLSLEQLHLLQNTREALGVLTDLDSGDSPEYALHPS
PLGEGCKRQNRVVPFTYQSLKQVLSLILNDVDFEYFKRLNSVYFCLPAGELALGE
TYAEILGRVGSPIPADAYVSSVPCASCFAEAMLPNGESVLSMLAAVNCNHCVRQ
VDSAPFVGFENRHLRGADARAGGAGRESREGREDADDEDEEPDRDQGDCA
GDSAAALRVLTSESSLVLAHGTIFEEEDGRLAELSNLYWSSNADRGVGGGTARATSSS

```

Job time: 1194 sec

polya_signal
CDS
/note="orf E3, 03, E4 and 06"
32195 34819
/note="orf 08"
/codon_start=1
/product="glycoprotein P"
/protein_id="AAC13795.1"
/db_xref="GI:695180"
/translation="MGVGGPRVVLWCVAALLCGVAQEVVAETTFPFATHREPVV
AENPAPFLPFCVCGASPTGGEIPEPELESCPTNEDKHIEGIALIYKINIVPVF
NVRKYIMTSITIKGSEDAIINQHTSYAVPLIEVQMDHYQCFSAVQVNEG
VNTYDRGHNETAELKPADGLTSITRYQSQPEVATPRNLWSYTRTYVCEVTE
MSARSMAPFEFFVTSVGDTEIEMSPFLAENGTEPEKILAPRSIQLLKNTAVTKYVGL
QADNATREFAIFGDSLSWKATTENSSYCDLILWKFSNAIQTHNSSLHFIAINDIT
ASFSTPLEEANFETKCIWNNTQEEIQKLEKEVETHRPNTAKYVTKTGNLYIY
QLIQIDLTHAKLYLNATASPTSTPTSPRRRRTDSSVSGGNGDNDSTKEES
VAAQVQFAYDNLKRSINRVLGELSRACREYRASLWMLSKINPTSVASAIYGP
VSAKLIGDVVSDGISVDQKSVFHNKMKVPGKEDLCYTPVGVGFNGFELFAGQ
LGPNEVLSTQVEVCOHSCHEVFQNGQYKYDYIYSTNLADIDIFLHMTLN
LSLVNIDFVLELYSKTEKELSNVDEIEMREKYNITQNLGRLKDDSDIDGRD
SFQITGLDLMQDLGLIGRVYVNVASGVSFGSIVSGVISFFKNPFGMLLIVLIAG
VVVYLPMHRSIISAPRIYIPGVERAAQEPGAHPVSEQIRNLGMHQFOQOR
AEEARREEVAGKRTLEFVIRDSATSVLRRRRGGGQYRQLRDGDSDDEYELRQ
DGYDDVDVEAGTADTV"
34855. 34860
/note="orf 07 and 08"
35089. 38115
/note="orf 09"
/codon_start=1
/product="DNA polymerase replicative subunit"
/protein_id="AAC13796.1"
/db_xref="GI:695181"
/translation="MSFYNPLVYKRTFLKAAPSRTKEYTRIIPKCFKTPGAAGVVP
RTSLDPVFGDKETPLIYGDGSRSLAGRGPGTGAGQCHTPVALTFHYVDIE
TVYGDQCHVFPQTDIIIPSGTVLKLGRISDDRSVNVNPROELVYFVYVPEGLK
LDLFIQQCSRENFPSQRYEYKTSKRYEYVYEAAREVYVYVSSQGFVDLLAGL
TAAGCEVFETNDAAARFIIDNGFTFGWYSCAAVPROGGAARUSWLEYDCAAGL
LEFHAGRADPGYVNLSEDFECIGENGFPNARSDMLQISCIWAKSGSEAPRSVL
NLNLTKEFTRVRSASSIEYHKKFNSSAGFMKAVSKVAGVVPIDMYQVCRDKLSL
NYKLDYVAGECVGAKEDEVSYKEIPELFRQGGARLGLCYKDSALVLDLLRYMT
HYEISEIKAKIPTRRVLTGQQINVFSCLLDVAGREGYILFVDRHADAGYOGATV
IDPSGFYNTPLVYFASLYPTIIQAHNLACYSTMIPGDRCLHPHLPGLGPDYEFELA
SGPVHEVKKHVASLALTNLWAKRAIRRELATVSDAETHILDKQOLAIVTCN
AVYFTGVASGILFCLKIAETVTFQGRMLNSKRYIEGVTPGLADILGRVECAPD
ASEKVIYGDTSLSLTHCGRYRPEQVTGFCDELAHMTRTLFDVPIKLEAKTFKLLIL

Query Match	13.8%	Score 32.4	DB 16	Length 184427
Best Local Similarity	48.9%	Pred. No. 13		
Matches	87	Conservative	0	Mismatches 91
				Indels 0
				Gaps 0
Qy	12	GCTGCCCCCTTCCTCTTTCGACCTTCCTTGGACGCTCACATGACAGCGCGGGGTAT	71	
Dy	85876	GGCTCCCCCGATCTCATCAACTCTGCTCACTCTTCTGTCACCTGTCGCGCCT	86817	
Qy	72	GACTTTGCAACTGAAGTGAAGTCTTTCTGACAAATTCCTCATGATCAGCTT	131	
Dy	86816	GGTTCGGCTCTCGGATGACCTGCTTCCCCCACCAGCCCTCTACGAGCCCATCA	86757	
Qy	132	CCTGGATTGCTGAAGTCTGCTCTCTCTCCATCTCCCTTACAGGACACAGCTC	189	
Dy	86755	GCTGAAGAACATCAAGCAGCGGGTGAAGCTGCAGTCCCCGAGCGGCCAGCATC	86599	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query	Score	Query			ID	Description
			Match	Length	DB		
1		234	100.0	439	1	V84366	Human stomach carcinoma
2		30	12.8	7824	1	V74052	Human interleukin-1
3	C	30	12.8	14690	1	X22303	Human IL-1ra BAC clone
4	C	29.8	12.7	291	1	V83065	EST clone AR34. Ne
5	C	29.4	12.6	6175	1	T34291	Coding sequence for
6	C	29	12.4	3580	1	V81742	Human p7p04 encodi
7	C	28	12.0	5018	1	N50150	Clone p2b6p102 enc
8	C	28	12.0	5018	1	N50150	Sequence of the CD
9	C	28	12.0	5962	1	Q70729	TATA-binding prote
10	C	28	12.0	5962	1	T43215	Human TATA-binding
11	C	28	12.0	5962	1	T79506	TATA-binding prote
12	C	27.8	11.9	2782	1	V71227	CNA encoding a mu
13	C	27.6	11.8	2635	1	T95763	Arabidopsis SCAREC
14	C	27.6	11.8	1853	1	V06751	Dentatorubral-pall
15	C	27.6	11.8	1293	1	V61832	Coding sequence fo
16	C	27.6	11.8	1697	1	V61833	Coding sequence fo
17	C	27.4	11.7	2135	1	T87016	Human FC-gamma rec
18	C	27.4	11.7	10023	1	V52731	Human hepatocyte n
19	C	27.4	11.7	984	1	V52663	Human native hepat
20	C	27.2	11.6	2121	1	T12282	Mouse TRAF2 cDNA.
21	C	27	11.5	5287	1	Q25440	E. coli DNA polyme
22	C	27	11.5	2459	1	Q85437	Actual shortened B
23	C	27	11.5	2652	1	T08977	Actual shortened B
24	C	27	11.5	1692	1	T60974	NLRK2 cDNA 3' reg
25	C	27	11.5	1785	1	T66986	Human brain-specif
26	C	26.8	11.5	2259	1	Q05106	Sequence encoding
27	C	26.8	11.5	2306	1	Q05873	Sequence encoding
28	C	26.8	11.5	3979	1	Q06312	Sequence of plasm
29	C	26.8	11.5	1083	1	Q12910	VNTR locus D6S22 f
30	C	26.8	11.5	1148	1	T28846	Mouse ccsackievir
31	C	26.8	11.5	11624	1	V52850	Human eyal gene co
32	C	26.6	11.4	999	1	Q12870	Promoter sequence
33	C	26.6	11.4	809	1	V37133	DNA sequence used
34	C	26.4	11.3	2483	1	Q14425	TK negative tkns 9
35	C	26.4	11.3	37913	1	Q51128	Excosepentaenoic a
36	C	26.4	11.3	37895	1	T34137	Gene for biosynthe
37	C	26.4	11.3	1106	1	T79671	BRCA2 cancer susc
38	C	26.4	11.3	37895	1	V00350	S. putrefaciens ei
39	C	26.4	11.3	808	1	V28688	Ripening banana bu

QY	1	ATGGGCTGTGGCTGCCCTTGTCCCTCTTGTACCCCTCCTTGGCAGCTCATATGGACA	60
Db	75	ATGGGCTGTGGCTGCCCTTGTCCCTCTTGTACCCCTCCTTGGCAGCTCATATGGACA	134
QY	61	GGGCGGGTATGACTTTGCACTGGAAGCTGAAGGAGCTTTTCTGACAAATTCCTCTAT	120
Db	135	GGGCGGGTATGACTTTGCACTGGAAGCTGAAGGAGCTTTTCTGACAAATTCCTCTAT	194
QY	121	GAGTCCAGCTCTCTGGAAATGCTTGAAGCTCTGCTCCTCTCCATCTCCCTTCAGGG	180
Db	195	GAGTCCAGCTCTCTGGAAATGCTTGAAGCTCTGCTCCTCTCCATCTCCCTTCAGGG	254
QY	181	ACGAGGGTACCCCTCCACCATGCAAGATCTCAACACCATGTGTCTGCAACACA	234

```
Db 255 ACCAGGTCACCCCTCCACCATGACAGATCTACACCATGTTGTCTGCAACACA 308
|||||
RESULT 2
ID 074052 standard; DNA: 7824 BP.
AC 074052:
DE 29-JAN-1996 (first entry)
DE Human interleukin-1-beta
KW Interleukin 1 beta; primer; mRNA; specificity; pharmaceutical: ss.
OS Homo sapiens. Location/Qualifiers
FH Key misc_difference 1331
FT /*tag= a
FT /note= "N is undefined"
FT J07123984-A.
PD 16-MAY-1995.
PD 05-NOV-1993: 275852.
PR 05-NOV-1993: JP-275852.
PA (HITB ) HITACHI CHEM CO LTD.
DR WPI; 95-211627/26.
PT A primer for the detection and the determ. of a specific messenger
PT RNA - can detect and determine specific mRNA(s) with high
PT reliability
PS Example 18; Page 17-20; 35pp; Japanese.
CC 074052 is interleukin-1-beta cDNA and 074019-21 are primers used for
CC the amplification of this cDNA. They are used specifically for the
CC detection and isolation of this sequence. The primers have the
CC advantage of high sensitivity and reliability and are useful in the
CC pharmaceutical industry.
CC Sequence 7824 BP; 2099 A; 1905 C; 1624 G; 2195 T;
SQ

Query Match 12.8%; Score 30; DB 1; Length 7824;
Best Local Similarity 55.9%; Pred. No. 5.6;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 73 ACTTTGCACTGAGCTGAGGAGCTTTCTGACAAATTCCTCTATGAGTCGAGCTTC 132
Db 5839 ACTCTACAGCTGGAGGTAAAGTGAATGCTATGGAATGAGCCCTTCTCAGCTTCGCTAC 4284
QY 133 CTGGAATTGCTTGAAGAAGCTCTGCTCTCTCTCCATCTCCCT 174
Db 4283 CACTTATTTCCGACAAACACCTTCTCTCCGCCGCCATCCCT 4242

RESULT 4
V86365/c
ID V86365 standard; cDNA: 291 BP.
AC V86365;
DT 27-APR-1999 (first entry)
DE EST clone AR34.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
KW Homo sapiens.
OS WO9845435-A2.
FN 15-OCT-1998.
PD 10-APR-1998; U06954.
PR 10-APR-1997; US-835913.
PA (GEM ) GENETICS INST INC.
PI Acostino M, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070076/05.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 210; 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC therapy. The EST sequences are also stated to be useful for gene
SQ Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;

Query Match 12.7%; Score 29.8; DB 1; Length 291;
Best Local Similarity 70.2%; Pred. No. 1.9;
Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 69 TATGACTTGCACGTGAGCTGAGGAGCTTTTCTGACAAATTCCTCTATGAGTC 125
Db 257 TTGATTTTGTACAAAGCTGAGTATTTTCTTACACATTTCTCTTAAATC 201

member of the cytokine superfamily that is expected to inhibit
inflammation by binding to the interleukin-1 receptor (IL-1R). It may
also bind to a new receptor so could regulate other cellular processes
associated with acute or chronic inflammation, e.g. asthma, chronic
myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
bowel disease. It may also induce or suppress interleukins, cytokines and
growth factors. Modulators of this protein are used to treat or prevent
conditions associated with abnormal levels of inflammation, or activity
of IL-1 or its receptor complex.
Sequence 14690 BP; 3810 A; 3245 C; 3299 G; 4327 T;
SQ
```

RESULT 5
T34291/c
ID T34291 standard; cDNA to mRNA; 6175 BP.
AC T34291;
DT 25-MAR-1997 (first entry)
DE Coding sequence for smooth muscle myosin heavy chain SM1.
KW Smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;
KW gene therapy; mouse; SM2 isoform; retrovirus; adenovirus; restenosis;
KW associated adenovirus; coronary artery catheterisation; sclerotic artery;
KW ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 105..6023
FT /*tag= a
FT /product= SM1
FN WO9623069-A1.
PD 01-AUG-1996.
PF 25-JAN-1995; J00134.
PR 25-JAN-1995; JP-010085.
PA (OSAP) OSAKA PREFECTURAL GOVERNMENT.
PY (VSS-) VESSEL RES LAB CO LTD.
PI Araiawa E, Hasegawa K, Ishiyama H, Matsuda Y, Oda S;
FI Sugawara M, Takahashi K.
DR WPI; 96-362693/36.
DR P-PSDB; W00024.
FT DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in
FT a vector for gene therapy for reduction of re-stenosis following
FT coronary artery catheterisation
PS Claim 5; Page 14-27; 42pp; Japanese.
CC This sequence represents the coding sequence for the smooth muscle myosin
CC heavy chain SM1 isoform protein. This SM1 coding sequence was isolated
CC from a mouse embryo cDNA library using probes based on the rabbit smooth
CC muscle myosin heavy chain SM2 isoform. This sequence is included in the
CC gene therapy vector of the invention. The vector of the invention
CC consists of this sequence inserted into a retrovirus, adenovirus,
CC associated adenovirus or animal-expression plasmid vector (such as pCMX2
CC or pAG208). The vector can be used in the gene therapy treatment of
CC arteriosclerosis, particularly for the reduction of restenosis occurring
CC following coronary artery catheterisation for widening of sclerotic
CC arteries.
SQ Sequence 6175 BP; 1846 A; 1455 C; 1856 G; 1018 T;

Query Match 12.6%; Score 29.4; DB 1; Length 6175;
Best Local Similarity 54.1%; Pred. No. 7.9;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 124 TCCAGCTTCTCGATTCGTTGAAGAGCTGCTCCCTCCCTCCATCTCCCTTCAGGGACC 183
DB 302 TCCATTTCACCAACTCCAGCACCTCATCGCCCTTCCTCTCTGATGCTGGCTGC 243
QY 184 AGCTGACCTCCACATGCAAGATCTCAACACCATGTTGCTGCAACACA 234
DB 242 TTCGAAGCCCTGCTTCTGGAAGGACCCACACAGCTTCTGGCTACCCA 192

RESULT 6
V81742/c
ID V81742 standard; cDNA; 3580 BP.
AC V81742;
DT 10-MAR-1999 (first entry)
DE Human PTP04 encoding cDNA.
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 53..2476
FT /*tag= a
FN WO9849317-A2.
PD 05-NOV-1998..

PF 27-APR-1998; U08439.
PR 23-OCT-1997; US-063595.
PR 28-APR-1997; US-044428.
PR 20-MAY-1997; US-047222.
PR 11-JUN-1997; US-049477.
PR 11-JUN-1997; US-049756.
PR 18-JUN-1997; US-049914.
PA (SUGEN-) SUGEN INC.
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,
PI Markby D, Orrust S, Peles E, Plowman GD;
PI WPI; 99-009434/01.
DR P-PSDB; W89247.
DT New nucleic acid encoding specific protein tyrosine phosphatases -
FT useful for identifying specific modulators for treatment and
FT prevention of cancer and neurodegenerative disease
PS Claim 2; Page 145-146; 193pp; English.
CC The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes human PTP04. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival.
CC Particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
SQ Sequence 3580 BP; 1213 A; 666 C; 618 G; 1083 T;

Query Match 12.4%; Score 29; DB 1; Length 3580;
Best Local Similarity 48.0%; Pred. No. 8.6;
Matches 83; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 23 TCCTCTCTTGACCTCTCTGGAGCTCACATGGAACAGGCGGGGTATGACTTTGCAAC 82
DB 642 TCATGCTCTGGCCAAATCTTGTATGAACCTGGTAGATAGTTCGAGTTTCACTATTGAAC 583
QY 83 TGAAGCTGAAGAGAGTCTTCTTGACAAATTCCTCTATGATCCAGCTTCCTGGAATGC 142
DB 582 TTAACCTTTAGAGTCTCGATTATATATACAGATTTCTTTTTCAGCTTCACAGATACA 523
QY 143 TTGAAGAGCTCTGCTCTCTCTCTCCATCTCCCTTCAGGGACCGCTCACCTC 195
DB 522 GAGAAAGGCCAANAATCCAGCTGCATCTCTCTCTGCTCAGCCAGTAGCGCTC 470

RESULT 7
N70646/c
ID N70646 standard; cDNA; 5018 BP.
AC N70646;
DT 27-APR-1991 (first entry)
DE Clone 23B6p102 encoding polypeptide with IgE binding factor
DE activity.
KW Antibody; immunoglobulin; ss.
FH Key Location/Qualifiers
FT CDS 423..2753
FT /*tag= a
FN J62045600-A.
PD 27-FEB-1987.
PF 21-AUG-1985; 183810.
PR 21-AUG-1985; JP-183810.
PA (SCHE) SCHERING BIOTECH CO.
DR WPI; 87-097765/14.
DR P-PSDB; P70417.
FT cDNA clone coding polypeptide - showing IgE bond factor activity.
PS Disclosure; Page 29; 33pp; Japanese.
CC The clone encodes a polypeptide which shows IgE binding factor

CC activity. See also N70645.
SQ Sequence 5018 BP; 1423 A; 1087 C; 1165 G; 1343 T;

Query Match 12.0%; Score 28; DB 1; Length 5018;
Best Local Similarity 53.7%; Pred. No. 20;
Matches 58; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 116 CCTATGAGTCCAGCTTCCTGGAATGCTTGAAAGCTCTGCCTCCCTCCATCTCCCTT 175
DB 3138 CATCTGACGGCAGAACTGCTGAAATTTTGGACGATATAGCTGGTCCATATCAGTCTT 3079

QY 176 CAGGACACAGCGTCCACCTCCACATGCAAGATCTCAACACCATGTTG 223
DB 3078 AAGGAGTCTGGGTTTCCCAAGCACTCCATGCTCCCAAGCAATGTTG 3031

RESULT 8
N50150/c
ID N50150 standard; cDNA; 5018 BP.
AC N50150;
DT 27-SEP-1991 (first entry)
DE Sequence of the cDNA clone 23B6p10.2 encoding a polypeptide
DE exhibiting mammalian immunoglobulin binding factor activity (IBF).
KW Immunoglobulin E-mediated disease; therapy; B-cell differentiation;
KW Immunoglobulin E-binding factor; ss.
PH Key Location/Qualifiers
FT CDS
EP-155192-A.
PD 18-SEP-1985.
PF 15-MAR-1985; 301834.
PR 16-MAR-1984; US-590430.
PA (SCHE-) SCHERING BIOTECH CO.
PA (UYJO) JOHNS HOPKINS UNIV.
PA (SCHE) SCHERING-BIOTECH CO.
PA (DNAX-) DNAX RES INST MOLEC.
PI Martens CL, Ishizaka K, Moore KW, Huff TE;
PI WPI; 85-231863/38.
DR P-PSDB; P50121.
DT New complementary DNA clones coding for poly(peptide(s)) - with
DT sequence of mammalian immunoglobulin factor and obtd. from
PT transformed or transfected host
PT Claim 8; Page 59-64; 71pp; English.
CC IBF and IBF cDNA are useful in studies on the immune system.
CC Treatment for IgE-mediated diseases may be possible. IBF may enhance
CC B-cell differentiation into an immunoglobulin-secreting cell.
SQ Sequence 5018 BP; 1424 A; 1086 C; 1165 G; 1343 T;

Query Match 12.0%; Score 28; DB 1; Length 5018;
Best Local Similarity 53.7%; Pred. No. 20;
Matches 58; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 116 CCTATGAGTCCAGCTTCCTGGAATGCTTGAAAGCTCTGCCTCCCTCCATCTCCCTT 175
DB 3138 CATCTGACGGCAGAACTGCTGAAATTTTGGACGATATAGCTGGTCCATATCAGTCTT 3079

QY 176 CAGGACACAGCGTCCACCTCCACATGCAAGATCTCAACACCATGTTG 223
DB 3078 AAGGAGTCTGGGTTTCCCAAGCACTCCATGCTCCCAAGCAATGTTG 3031

RESULT 9
Q70729/c
ID Q70729 standard; cDNA; 5962 BP.
AC Q70729;
DT 23-MAR-1995 (first entry)
DE TATA-binding protein-associated factor hTAFII250 cDNA.
DE TATA-binding protein associated factor; hTAFII250; ss; screening;
KW diagnostic; therapeutic; gene transcription regulation.
OS Homo sapiens. Location/Qualifiers
FH Key

FT cds 14..5692
PN WO9417087-A. /*tag= a
PD 04-AUG-1994.
PF 28-JAN-1994; U01114.
PR 28-JAN-1993; US-013412.
PR 30-JUN-1993; US-087119.
PA (REGC) UNIV CALIFORNIA.
PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
PI WPI; 94-264019/32.
DR P-PSDB; R56491.
DT TATA-binding protein associated protein factors - and
DT corresponding nucleotide sequence and deriv. antibodies, useful
DT in screening diagnostics and therapeutics
PS Disclosure; Page 94-112; 180pp; English.
CC The TATA-binding protein associated factor hTAFII250 (including
CC specific antibodies and fusion products) are used in drug screening,
CC diagnostics and therapeutics. They are used in the development of
CC specific biochemical assays for screening compounds that agonise or
CC antagonise selected transcription factors involved in regulating
CC gene expression associated with human pathology.
SQ Sequence 5962 BP; 1764 A; 1294 C; 1501 G; 1403 T;

Query Match 12.0%; Score 28; DB 1; Length 5962;
Best Local Similarity 48.2%; Pred. No. 21;
Matches 79; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 11 GGCTGGCCCTTGTCTCTCTGTGACCCCTCTTGGCAGCTCACATGGAACAGGGCGGTA 70
DB 3993 GTCTTTTCCAACTCTCTCTCTGTCTCTGTCTGTGTCGACAGGGTTGGAAGGTGGCGCA 3934

QY 71 TGACTTTGCAACTGAGCTGAGGAGCTCTTTCTGACAAATCTCTATGAGTCCAGCT 130
DB 3933 TTGCTTTGATAATAGAGGGGCGAGAAFTTGTAGTCTCATGTGTCATGCGCACCAT 3874

QY 131 TCTTGGAAATGCTTGAAAGCTGTGCTCTCTCTCCATCTCCCT 174
DB 3873 GCCCCACATTTTCAGTTTTAGTTCAGGACGCTCTCTTCATTTTCTT 3830

RESULT 10
T42215/c
ID T42215 standard; cDNA; 5962 BP.
AC T42215;
DT 27-JAN-1997 (first entry)
DE Human TATA-binding protein associated factor hTAFII250 gene.
DE Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
KW lambda-gt11; expression library; ds.
OS Homo sapiens. Location/Qualifiers
FH Key 14..5695
FT CDS /*tag= a
FT FT /product= human TAFII250
PN US5534410-A.
PD 09-JUL-1996.
PF 28-JAN-1993; 013412.
PR 28-JAN-1993; US-013412.
PR 30-JUN-1993; US-087119.
PR 28-JAN-1994; US-188582.
PA (REGC) UNIV CALIFORNIA.
PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
PI WPI; 96-333245/33.
DR P-PSDB; W06082.
DT Screen for cDps. that bind human TATA-binding protein associated
DT factor - by testing ability to bind to polypeptide fragments of the
DT factor, useful as (ant)agonists of transcription factors involved in
DT disease.
PS Examples; Column 65-78; 86pp; English.
CC This is the nucleotide sequence encoding the human TATA-binding protein

OS Homo sapiens.
PD US5723301-A.
PF 03-MAR-1998.
PR 03-NOV-1995; 553110.
PT (UYDU-) UNIV DUKU.
PI Burke JR, Enghild J, Strittmatter WJ, Vance JW;
PT WPI; 98-178475/16.
PT Screening assay for inhibitors of GAPDH binding to polyglutamine -
PT for use in treating neuro-degenerative diseases or psychiatric
PT disorders

PS Example 7; Columns 21-24; 19pp; English.
PS This sequence represents dentatorubral-pallidoluysian atrophy (DRPLA)
CC cDNA (nucleotides 238-2090 of Genbank accession number D38529). The
CC invention relates to a new screening assay for compounds capable of
CC inhibiting binding of glyceraldehyde-3-phosphate dehydrogenase (GAPDH)
CC to polyglutamine. The assay comprises providing an aqueous solution
CC containing a test compound, polyglutamine-containing molecules and GAPDH
CC and determining if binding of GAPDH to the polyglutamine-containing
CC molecules is reduced in the presence of the test compound. Inhibitors
CC identified by the above assay could be used to treat neurodegenerative
CC diseases or psychiatric disorders caused by a gene containing an
CC extended CAG domain, e.g. Huntington's disease or schizophrenia.
CC Sequence 1853 BP; 347 A; 774 C; 388 G; 344 T;

SQ Query Match 11.8%; Score 27.6; DB 1; Length 1853;
Best Local Similarity 56.7%; Pred. No. 18;
Matches 51; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 112 TCCTCCTATGATCGACGTCTCTGGAAAGCTTCGCTTCTCTCCATCTC 171
DB 940 TCTTCTCTGCCTCCCGCTTCCCAGCATTCGCCAGCTACCCCCACTCTTC 999

OY 172 CCTTCAGGACGACGCTCACCCCTCCACCATT 201
DB 1000 CTCTCCCCAACAGCGCTCTCTGTCTCAAT 1029

RESULT 15
V61832/c
ID V61832 standard; cdNA; 1293 BP.
AC V61832;
DT 28-APR-1999 (first entry)
DE Coding sequence for human p53 regulated protein, P2XM.
KW Human; p53 regulated protein; P2XM; ATP receptor; RP-2 protein;
KW thymocyte; apoptosis; tumour-suppressor gene; Rhabdoid sarcoma;
KW anticancer drug; ss.
CS Homo sapiens.
PN M09842815-AI.
PD 01-OCT-1998.
PF 18-MAR-1998; J01146.
PR 26-MAR-1997; JP-03044.
PI (SAKA) OTSUKA PHARM CO LTD.
PA Nakamura Y, Tokino T;
DR WPI; 98-532006/45.
DR P-PDSB; W76434.

PT Human gene P2XM whose transcription is induced by p53 - useful, e.g.
PT for diagnostic purposes and in development of new anticancer drugs
PS Claim 2; Page 23-23; 43pp; Japanese.
CC This sequence represents the coding sequence for the P2XM protein of the
CC invention. The protein is significantly homologous to: (i) the P2X family
CC during apoptosis. Transcription of the genes is expressed in thymocytes
CC of ATP receptors, and (ii) RP-2 protein which is specifically regulated by
CC the tumour-suppressor gene p53. The P2XM gene is specifically expressed by
CC in skeletal muscle and has been localised to chromosome 22q11, an area
CC where mutation and sequence losses frequently occur in Rhabdoid sarcomas.
CC The genes may be used for diagnostic purposes (e.g. by detecting changes
CC occurring in the gene in sarcomas), using probes and primers containing
CC or derived from all or part of the genes. The genes may further be used
CC in the development of new anticancer drugs.

SQ Sequence 1293 BP; 271 A; 379 C; 389 G; 254 T;

Query Match 11.8%; Score 27.6; DB 1; Length 1293;
 Best Local Similarity 58.5%; Pred. No. 16;
 Matches 48; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 136 GAATTGCTTGAAGAAGCTCTGCTCTCTCCATCTCCCTTCAGGGACCAAGCGTCACCTTC 195
 ||| | |||| | |||| | || |||| |||| |||| |||| |||| ||||
 Db 567 GAACAGTGTGAAGTTCTGGGCTGGGCGACGAGGGGCTTCAGGGGCAACACCCACCTTC 508
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 196 CACCATGCAGATCTCAACCC 217
 ||| |||| |||| |||| ||||
 Db 507 CACGGGCGACCACTCCAGATC 486

Search completed: March 20, 2000, 09:31:54
 Job time: 1077 sec

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/172,528
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/937,466
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2782 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-172-528-1

Query Match 11.9%; Score 27.8; DB 4; Length 2782;
Best Local Similarity 53.2%; Pred. No. 5.7; Mismatches 0; Gaps 0;
Matches 59; Conservative 0; Indels 52; Gaps 0;
QY 56 GAACAGGCGGGTATGACTTGGCACTGAGCTGAAGAGTCTTTCTGACAAATTCCT 115
DB 2101 GAAACCTCGAGAGAGGACAGAGACGTCTTCTGCTTGGCCIAAGGTGGAG 2042
QY 116 CCTATGAGTCCAGCTCTCTGGAATGCTGAAAAGCTCTGCTCTCTCTCC 166
DB 2041 ATTCTGAGTCTGCTGCTTGAAGTATGATGAGCAGCAGTGCCTCCCTCC 1991

RESULT 7
US-08-553-110-2
Sequence 2, Application US/08553110
Patent No. 5723301
GENERAL INFORMATION:
APPLICANT: Burke, James R.
APPLICANT: Vance, Jeffrey M.
APPLICANT: Enghild, Jan
APPLICANT: Stittmatter, Warren J.
TITLE OF INVENTION: Therapeutics for Diseases Associated
TITLE OF INVENTION: with CAG Trinucleotide Repeat Expansion
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenneth D. Sibley
STREET: P.O. Drawer 34009
CITY: Charlotte
STATE: No. 5723301th Carolina
COUNTRY: United States of America
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,110
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sibley, Kenneth D.
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-117
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1853 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-553-110-2
Query Match 11.8%; Score 27.6; DB 2; Length 1853;
Best Local Similarity 56.7%; Pred. No. 5.5; Mismatches 0; Gaps 0;
Matches 51; Conservative 0; Indels 39; Gaps 0;
QY 112 TCCTCTATGAGTCCAGCTTCTGGAATGCTTGAAAGCTCTGCTCTCTCTCCATCTC 171
DB 940 TCTTCTCTGCTCCCTTCCAGCTTCCAGGCAATTCGCCAGCTACCCCCCATCTCTTC 999
QY 172 CCTTCAGGAGCAGCTCACCTCCACCAT 201
DB 1000 CTTCCCGACACAGCTCTCTGTCTCCCAT 1029

RESULT 8
US-08-331-394-3/c
Sequence 3, Application US/08331394
Patent No. 5670319
GENERAL INFORMATION:
APPLICANT: Goeddel, David V.
APPLICANT: Rothe, Mike
TITLE OF INVENTION: Tumor Necrosis Factor
TITLE OF INVENTION: Receptor-Associated Factors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/331,394
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/250858
FILING DATE: 27-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 897P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2121 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-331-394-3

Query Match 11.6%; Score 27.2; DB 1; Length 2121;
Best Local Similarity 53.8%; Pred. No. 8; Mismatches 0; Gaps 0;
Matches 56; Conservative 0; Indels 48; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2000, 08:46:58 ; Search time 538.4 Seconds
(without alignments)

1640.983 Million cell updates/sec

Title: US-09-092-296-7_COPY_51_284

Perfect score: 234

Sequence: 1 ATGGGGCTGGGCTGCCCT.....ACCATGTTCTGCAACACA 234

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST.*

Word size : 0

Number of hits that pass the threshold : 9077268

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*
- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*

- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*
- 54: em_est22.*
- 55: em_est23.*
- 56: em_est24.*
- 57: em_est25.*
- 58: em_est26.*
- 59: gb_est33.*
- 60: gb_est34.*
- 61: gb_est35.*
- 62: gb_est36.*
- 63: gb_est37.*
- 64: gb_est38.*
- 65: em_est27.*
- 66: em_est28.*
- 67: em_est29.*
- 68: em_est30.*
- 69: gb_est39.*
- 70: gb_est40.*
- 71: gb_est41.*
- 72: gb_est42.*
- 73: gb_est43.*
- 74: gb_est44.*
- 75: em_est31.*
- 76: em_est32.*
- 77: em_est33.*
- 78: em_est34.*
- 79: gb_gss1.*
- 80: gb_gss2.*
- 81: gb_gss3.*
- 82: gb_gss4.*
- 83: em_gss1.*
- 84: em_gss2.*
- 85: em_gss3.*
- 86: em_gss4.*
- 87: gb_gss5.*
- 88: gb_gss6.*
- 89: gb_gss7.*
- 90: gb_gss8.*
- 91: gb_gss9.*
- 92: em_gss5.*
- 93: em_gss6.*
- 94: em_gss7.*
- 95: em_gss8.*
- 96: em_gss9.*
- 97: em_gss10.*
- 98: em_gss11.*
- 99: gb_gss10.*
- 100: gb_gss11.*
- 101: em_gss12.*
- 102: gb_gss13.*
- 103: gb_gss13.*
- 104: gb_gss14.*
- 105: gb_gss15.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
C 1	175	74.8	404	61	AI857998		AI857998 wfj6b01.x
C 2	93.2	39.8	328	42	AI136523		AI136523 UI-R-C2p-
C 3	75.8	32.4	552	82	AQ718761		AQ718761 HS_5511_B
C 4	36.4	15.6	331	31	AA282040		AA282040 zs89a03-S
C 5	36.4	15.6	494	34	AA534569		AA534569 nf77c11.S
C 6	36.4	15.6	424	38	AA741185		AA741185 ob30a02.S
C 7	36.4	15.6	402	40	AA936790		AA936790 oh89a02.S

```

c 8 36.4 15.6 529 41 A1018775
c 9 34.8 14.9 451 61 A1829206
c 10 34 14.5 190 41 A1010074
c 11 33.4 14.3 545 23 T41524
c 12 33.2 14.2 292 40 A591822
c 13 33.2 14.2 597 63 A591822
c 14 32.6 13.9 522 82 A0718755
c 15 32.4 13.8 490 51 A1757641
c 16 32.2 13.8 875 82 A0740265
c 17 32 13.7 573 23 T41535
c 18 31.8 13.6 547 23 T41523
c 19 31.8 13.6 267 50 A062745
c 20 31.8 13.6 613 63 A1981365
c 21 31.8 13.6 568 100 A0290553
c 22 31.6 13.5 602 102 A0373909
c 23 31.4 13.4 505 20 M89011
c 24 31.4 13.4 433 23 R90433
c 25 31.4 13.4 523 23 T44883
c 26 31.4 13.4 548 23 T76643
c 27 31.4 13.4 529 37 H76456
c 28 31.4 13.4 497 37 T44613
c 29 31.2 13.3 615 59 A1765674
c 30 31.2 13.3 617 64 A1042852
c 31 31.2 13.3 637 64 A0253169
c 32 31 13.2 398 20 T52782
c 33 31 13.2 307 20 T45258
c 34 31 13.2 536 23 T41525
c 35 31 13.2 466 60 A1803834
c 36 30.8 13.2 459 46 A1442028
c 37 30.8 13.2 711 62 A076246
c 38 30.6 13.1 313 20 T36119
c 39 30.6 13.1 368 21 R15181
c 40 30.6 13.1 467 21 T59783
c 41 30.6 13.1 441 22 R34436
c 42 30.6 13.1 437 33 AA40444
c 43 30.6 13.1 257 33 AA443415
c 44 30.6 13.1 363 43 A1219495
c 45 30.6 13.1 471 50 A1692905

```

ALIGNMENTS

```

RESULT 1
LOCUS A1857998/c 404 bp mRNA EST 26-AUG-1999
DEFINITION wj69b01.x1 NCI_CGAP_Lu19 Homo sapiens CDNA clone IMAGE:2408041 3',
mRNA sequence.
ACCESSION A1857998
VERSION A1857998.1 GI:5511614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor gene index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```

```

Seq primer: -40UP from Gibco
High quality sequence stop: 395.
FEATURES
source
1..404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 97 a 105 c 117 g 84 t 1 others
ORIGIN
Query Match 74.8%; Score 175; DB 61; Length 404;
Best Local Similarity 99.4%; Pred. No. 6.6e-41;
Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 59 CAGGCCGGGTATGACTTCGAAGCTGAGGAGTCTTTCTGACAAATCCCTCT 118
Db 311 CAGGCCGGGTATGACTTCGAAGCTGAGGAGTCTTTCTGACAAATCCCTCT 252
QY 119 ATGATGCCAGCTTCCTGAATTCGTTGAAAAGCTCTCCCTCCATCCCTCAG 178
Db 251 ATGATGCCAGCTTCCTGAATTCGTTGAAAAGCTCTCCCTCCATCCCTCAG 192
QY 179 GGACACGCTGACCTCCACCATGCAAGATCTCAACACCATGTTGTCTCAACACA 234
Db 191 GGACACGCTGACCTCCACCATGCAAGATCTCAACACCATGTTGTCTCAACACA 136

```

RESULT 2

```

A1136523/c
LOCUS A1136523 328 bp mRNA EST 11-FEB-1999
DEFINITION UT-R-C2p-nq-e-02-0-UI.s1 UT-R-C2p Rattus norvegicus cDNA clone
UT-R-C2p-nq-e-02-0-UI 3', mRNA sequence.
ACCESSION A1136523
VERSION A1136523.1 GI:3637300
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150222.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
431 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iue.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult lung library. cDNA Library preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research

```


High quality sequence stop: 242.

FEATURES

source
1. .331
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

97 a 61 c 51 g 122 t

Query Match 15.6%; Score 36.4; DB 31; Length 331;
Best Local Similarity 56.8%; Pred. No. 1.2;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 87 GCTGAAGAGCCTTCCTGACAAATTCCTCTATGAGTCCAGCTTCCTGGAATTCCTTGA 146
|||||
Db 318 GATGATGACTAGTCTGATATATAATCCACTAAATCCAGTACCAGGAACCTGCCTGG 259
|||||
QY 147 AAAGCTCTGCTCCTCCATCTCCCTTCAGGACCGAGTCCACCTCCACCATGCA 204
|||||
Db 258 AACTGTGGCCATGCTATTTTCTTTAAAGACCAAGTGATAGTAGGCCATGCA 201
|||||

RESULT 5
AA534569/c

LOCUS nf77c11.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:925940 3',
DEFINITION mRNA sequence.
ACCESSION AA534569
VERSION AA534569.1 GI:2278822
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802984.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/URL at: www.bio.llni.gov/bdrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 370.
Location/Qualifiers

source

1. .494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization."

BASE COUNT 139 a 108 c 94 g 153 t
ORIGIN

Query Match 15.6%; Score 36.4; DB 34; Length 494;
Best Local Similarity 56.8%; Pred. No. 1.3;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 87 GCTGAAGAGCCTTCCTGACAAATTCCTCTATGAGTCCAGCTTCCTGGAATTCCTTGA 146
|||||
Db 296 GATGATGACTAGTCTGATATATAATCCACTAAATCCAGTACCAGGAACCTGCCTGG 237
|||||
QY 147 AAAGCTCTGCTCCTCCATCTCCCTTCAGGACCGAGTCCACCTCCACCATGCA 204
|||||
Db 236 AACTGTGGCCATGCTATTTTCTTTAAAGACCAAGTGATAGTAGGCCATGCA 179
|||||

RESULT 6
AA741185/c

LOCUS Ob30a02.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1325162 3',
DEFINITION mRNA sequence.
ACCESSION AA741185
VERSION AA741185.1 GI:2779777
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 5, 1998 this sequence version replaced gi:2581148.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
unknown library type
Insert length: 742 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 387.
Location/Qualifiers

FEATURES
source

1. .424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Kid5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-AACTGGAAGATCGCGCGCAATTTTTTTTTTTTTTTT-3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and Eco RI sites of the modified pT73 vector. Library

87 GCTGAGGAGTCTTTCTGACAAATTCCTCATGAGTCCAGCTTCCTGGAATTCCTTGA 146

Db 299 GATGATGCTACTGCTCATATATTAATCCACTTAATCCAGTACCTACAGGACCTGCTGG 240

QY 147 AAAGTCGCTCCTCTCTCCATCTCCCTTTCAGGACGACGCTCACCCTCCACCATGCA 204

Db 239 AACTGTGGCCATGCAATTTTTTTTTTTCTTAAGACCACTGTGTAGTAGGCGCATGCA 182

RESULT 9

AI829206/c 451 bp mRNA EST 26-AUG-1999

LOCUS WK76101.x1 NCI_CGAP_Panl Homo sapiens cDNA clone IMAGE:2421337 3', mRNA sequence.

DEFINITION AI829206

ACCESSION AI829206

VERSION 1 GI:5449877

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 451)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3189627. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Life Technologies catalog #: 11548-013 DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/obrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 426.

FEATURES

source

1..451

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2421337"

/clone_lib="NCI_CGAP_Panl"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

BASE COUNT 127 a 97 c 85 g 142 t

ORIGIN

Query Match 14.9%; Score 34.8; DB 61; Length 451;

Best Local Similarity 55.9%; Pred. No. 3.8;

Matches 56; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 87 GCTGAAGAGCTTTTCTGACAAATTCCTTCATAGTCGACGCTCCAGCTTCCTGCAATTCCTTGA 146

Db 297 GATGATGCTACTGTACTGTATATAATCCACTTAATCCAGTACCTACAGGAACTCCCTGG 238

QY 147 AAAGTCGCTCCTCTCCATCTCCCTTCAGGACGACGCTCACCCTCCACCATGCA 204

Db 237 ACCTGTGGCATGCATTTTTTTTTTTCTTAAGACCACTGTGTAGTAGGCGCATGCA 180

RESULT 10

AI010074/c 190 bp mRNA EST 15-JUN-1998

LOCUS EST204525 Normalized rat lung, Bento Soares Rattus sp. cDNA clone

DEFINITION RI0BT38 3' end, mRNA sequence.

ACCESSION AI010074

VERSION AI010074

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 190)

AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus. Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D. Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index

TITLE Unpublished (1998)

JOURNAL On Jan 19, 1998 this sequence version replaced gi:2150555.

COMMENT Contact: Lee, NH ATCC The Institute for Genomic Research 9712, Medical Center Drive, Rockville, MD 20850, USA Tel: (301)-838-3529 Fax: (301)-838-0208 Email: nhlee@tigr.org Seq primer: M13-21.

FEATURES

Location/Qualifiers

1..190

Source

/organism="Rattus sp."

/db_xref="taxon:10118"

/clone="RI0BT38"

/clone_lib="Normalized rat lung, Bento Soares"

/note="Organ: lung; Vector: p1773Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 39 a 51 c 59 g 41 t

ORIGIN

Query Match 14.5%; Score 34; DB 41; Length 190;

Best Local Similarity 59.7%; Pred. No. 4.7;

Matches 46; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 164 TCCATCTCCCTTCAGGACGACGCTCACCCTCCACCATGCAAGATCTCACACCATGTGG 223

Db 190 TCCACCTCTCACCGGGGACCAATGTACCCCTTCATCATTAAGGGCCACACCATCTTA 131

QY 224 TCTGCA 229

Db 130 CTTGCA 125

RESULT 11

T41524 545 bp mRNA EST 07-AUG-1995

LOCUS 10105 Lambda-PRL2 Arabidopsis thaliana cDNA clone 100A7T7, mRNA sequence.

ACCESSION T41524

VERSION T41524

KEYWORDS EST.

SOURCE Arabidopsis thaliana

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 545)

AUTHORS Newman, T., deBuijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M., Retzel, E. and Somerville, C. Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1255 (1994)

TITLE Unpublished (1994)

JOURNAL On Jan 24, 1995 this sequence version replaced gi:634112.

COMMENT Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan State University MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI Tel: 517-353-0854

[illegible]

Wed Mar 22 12:07:56 2000

Db 128 TGCTGCTGTTGGACATCTTCCTTCAGCTGCTGGATACGCTCGGCAATCGCTGCGCTC 69
Oy 172 CCTTCAGGACCGGTCACCTCCACCATGCAAGATCTCAACACCATGTTGTTGTCGAC 231
Db 68 CCGTTTGGCCCCCGGACCTCCGCTCTTTGCTGCTATCTTTGCGTTCTTCGCTCGAC 9
Oy 232 AC 233
Db 8 GC 7

Search completed: March 19, 2000, 08:56:37
Job time: 579 sec



Wed Mar 22 12:07:35 2000

PT Prodn. of recombinant apo-lipoprotein E in insects - by infecting
PT Manduca sexta larvae with recombinant Autographica californica
PT nuclear polyhedrosis baculovirus vector.
PS Disclosure: Columns 11-14; 10pp; English.
CC Recombinant human apolipoprotein-E (ApoE) (R86791) can be produced
CC by preparing a genetic construct (contg. an ApoE-encoding sequence,
CC e.g. 106957, and flanking regulatory sequences enabling the protein
CC to be expressed in insect cells), which is then introduced into a
CC Manduca sexta larva (using a recombinant Autographica californica
CC nuclear polyhedrosis virus) and recovering the protein from the
CC haemolymph of the larval host. The ApoE produced is in a form
CC sufficiently complexed with lipids to be biologically active, which
CC cannot be achieved in insect cell cultures, and can therefore be
CC used in therapeutic applications.
CC Sequence 1157 BP; 212 A; 370 C; 426 G; 149 T;
SQ

Query Match 12.78; Score 29; DB 1; Length 1157;
Best Local Similarity 54.18; Pred. No. 4.3;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 6 GACTTCAGTGTCTCTCCATCCAGAGCGCCAGTGGCCACTATGGGGTCTGGGCTGCC 65
DB 626 GAGCGCGGCGCTCAGCGCCATCCGCGAGCGCTGGGGCCCTGTGTGAACAGGGCCGCTG 685
QY 66 TTGTCTCTCTTTGACCTCTCTTGGCAGCTCACATGGAACAGGGCCGGG 114
DB 686 CGGCGCGCCACTGTGGGCTCTCTGGCGCCAGCGGCTACAGGAGCGGG 734

Search completed: March 20, 2000, 18:37:54
Job time: 8528 sec

Bonaldo

Seq primer: -40UP from Gibco
High quality sequence stop: 395.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2408041"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I and oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 105 c 117 g 84 t 1 others
ORIGIN
Query Match 74.88; Score 175; DB 61; Length 404;
Best Local Similarity 99.48; Pred. No. 6.6e-41; Indels 0; Gaps 0;
Matches 175; Conservative 0; Mismatches 1;
QY 59 CAGGGCGGGTATCTTGAACCTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 118
Db 311 CAGGGCGGGTATCTTGAACCTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 252
QY 119 ATGAGTCCAGCTCTCTGGAATCTGGAAGCTGCTCCCTCCCTCCCTCCAG 178
Db 251 ATGAGTCCAGCTCTCTGGAATCTGGAAGCTGCTCCCTCCCTCCCTCCAG 192
QY 179 GSACACGCTGACCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCACACA 234
Db 191 GGACACGCTGACCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCACACA 136

RESULT 2
LOCUS A1136523/c 328 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C2p-nq-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
UI-R-C2p-nq-e-02-0-UI 3', mRNA sequence.
ACCESSION A1136523
VERSION A1136523.1 GI:3637300
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150222.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research

A1018775 ov32e12.x
A1829206 wk76f01.x
A1010074 EST204525
T41524 10105 Lambd
AA991822 ct45c07.s
AW013553 11P02sk W
A0718755 HS-5511.B
A1757641 ea33d07.y
A0740265 HS-5505.A
T41535 10116 Lambd
T41523 10104 Lambd
A0062745 A0062745
A1981365 pat.pk005
AQ290553 nbx0037C
AQ373909 RPI11-14
M89011 CEL14A7 Chr
R90433 16788 Lambd
T44883 8146 Lambd
T76643 11421 Lambd
H76456 18161 Lambd
T44613 7876 Lambd
A1765674 w183b02.x
A042852 DRFp434H
AW025169 wu94g07.x
T52782 va79f01.r1
Z45258 HSC2KH091.n
T41525 10106 Lambd
A1803834 t330d12.x
A1442028 sa66f12.y
A0076246 A0076246
T36119 EST96907 Hu
R15181 y86c11.r1
T59783 y86e04.r1
R34436 y957g06.r1
AA40444 zw37c01.r
AA43415 zw87d08.r
A1219495 qb24d03.x
A1692905 wd42f12.x

ALIGNMENTS

RESULT 1
LOCUS A1857998/c 404 bp mRNA EST 26-AUG-1999
DEFINITION wj69d01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408041 3',
mRNA sequence.
ACCESSION A1857998
VERSION A1857998.1 GI:5511614
KEYWORDS EST.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
RNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
obtained through the I.M.A.G.E. Consortium/LNL at:
http://lncat.nih.gov/bbrp/image/image.html

Genetics

Seq primer: M13 Forward.

Location/Qualifiers

FEATURES

source

1..328

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-C2p-ng-e-02-0-UI"

/clone_lib="UI-R-C2p"

/dev_stage="adult"

/lab_host="DHI08 (Life Technologies)"

/note="Vector: p773D-pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p

library is a subtracted library derived from the UI-R-C1

library, which is a subtracted library derived from the

UI-R-C0 library. The UI-R-C0 library consisted of a

mixture of individually tagged normalized libraries

constructed from rat placenta, adult lung, brain, liver,

kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day

embryo. The tag is a string of 3-5 nucleotides present

between the Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within

the mixture. The subtracted library (UI-R-C2p) was

constructed as follows: PCR amplified cDNA inserts from

UI-R-C1 clones from which 3' ESTs had been derived was

used as a driver in a hybridization with the UI-R-C1

library in the form of single-stranded circles. The

remaining single-stranded circles (subtracted library) was

purified by hydroxyapatite column chromatography,

converted to double-stranded circles and electroporated

into DHI08 bacteria (Life Technologies) to generate the

UI-R-C2p library. This procedure has been previously

described (Bonaldo, Lennon and Soares, Genome Research 6:

791-806, 1996)."

BASE COUNT 62 a 77 c 98 g 91 t

Query Match

Best Local Similarity 39.8%; Score 93.2; DB 42; Length 328;

Matches 128; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 44 GCAGCTCACATGGACAGCGGGGTATGACTTTGCAACTGAAGTGAAGGAGTCTTTTC 103

Db 327 GCAGCTCATCTGACAGGCGCAACAAAGTGACTTTGCAAGTGAAGTGAAGGAGTCTTTTC 116

QY 104 TGCAATTTCTCTCTATGAGTCCAGCTTCCTGGAAATGCTTGAAGAGTCTGCTCTCTCC 152

Db 267 AGCCCAAGACCTCCCAAGACTCGGGCTTTCTGGACATGCTCCAAAGATCTGCTCTCTCC 208

QY 164 TCCATCTCCTTTCAGGACGAGGTCACCTCCACATGCAAGATCTCAACACAGTGTG 223

Db 207 TCCATCTCTCAGCGGACCAATGTCACCTTTCATCAATAAGGCGCCACACACACCTTA 148

QY 224 TCTGCA 229

Db 147 CTTGCA 142

RESULT 3

A0718761/c

LOCUS

DEFINITION

HS_551132_F09_T7A RPI-11 Human Male BAC Library Homo sapiens

genomic clone Plate-1087 Col-18 Row-L, genomic survey sequence.

A0718761

ACCESSION

A0718761

VERSION

A0718761.1 GI:5468077

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallaceu.washington.edu

Clones are derived from the human BAC library RPI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)

or from Research Genetics (info@resgen.com). BAC end Web Server:

http://www.htso.washington.edu

Plate: 1087 row: L column: 18

Seq primer: T7

Class: BAC ends

High quality sequence stop: 552.

Location/Qualifiers

1..552

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-1087 Col-18 Row-L"

/clone_lib="RPI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Genomic sequence of BAC ends"

BASE COUNT 141 a 153 c 124 g 117 t 17 others

Query Match

Best Local Similarity 32.4%; Score 75.8; DB 82; Length 552;

Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 57 ACAGGGCGGGGTATGACTTTGCAACTGAAGTGAAGGAGTCTTTCTGACAAATTCCTC 116

Db 474 AGCATGGCGAGGTATGACTTTGCAACTGAAGTGAAGGAGTCTTTCTGACAAATTCCTC 415

QY 117 CTATGAGTCCAGCTTCTCGAATTCCTGAAAAGCT 152

Db 414 TTATGAGTCCAGCTTCTCGAATTCCTGAAAAGCT 379

RESULT 4

AA282040/c

LOCUS

DEFINITION

NCI-CGAP_81 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:704620 3',

mRNA sequence.

AA282040

ACCESSION

AA282040.1 GI:1924872

VERSION

AA282040.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1402231.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LBNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 891 Std Error: 0.00

Seq primer: -41m13 fwd. ET from Amerisham


```
##experimental_source strain AB972

GENETICS
#gene SGB:GCD6
##cross-references SGD:S0002619; MIPS:YDR211w
#map position 4R
KEYWORDS translation regulation
#length 712 #molecular-weight 81160 #checksum 142
SUMMARY

Query Match 16.5%; Score 92; DB 2; Length 712;
Best Local Similarity 39.5%; Pred. No. 1.23e+00;
Matches 17; Conservative 8; Mismatches 14; Indels 4; Gaps 4;

DBD 12 LGN-HGKNSMDVEDLQAVITDS-YETFRPLFAVPRCLL 52
      ||| : | : | : ||| : | : | : | : | : | : |
14 LGSSTGTGPTQLKLKESFUTNSSYESSFELLE-KL-CLL 54

RESULT 4
ENTRY type complete
TITLE osmotin-like protein precursor - tomato
ORGANISM #formal_name Lycopersicon esculentum #common_name tomato
DATE 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change
26-Aug-1999
JC5237 #status
JC5237 Chen, R.; Wang, F.; Smith, A.G.
#authors Gene (1996) 179:301-302
#journal A flower-specific gene encoding an osmotin-like protein from
#title Lycopersicon esculentum.
#cross-references M0ID:97128324
#contents flower
#accession JC5237
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-252 #label CHE
#cross-references GB:I76632; NID:q1220536; PID:q1220537
COMMENT This protein is rich in cysteine content and play a part of the
flowers defense mechanisms against pathogens.
CLASSIFICATION #superfamily thaumatocin I
FEATURE
1-23 #domain signal sequence #status predicted #label SIG\
24-252 #product Osmotin-like protein #status predicted #label
MAT
SUMMARY #length 252 #molecular-weight 27265 #checksum 2939

Query Match 16.3%; Score 91; DB 2; Length 252;
Best Local Similarity 54.2%; Pred No. 1.63e+00;
Matches 13; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Db 10 LPISLLFTLLSLSOSTNPFIILT 33
      ||| ||||| : ||| : | : |
Qy 5 LPPIVIMTMSCSHSGWPGKATLA 28

RESULT 5
ENTRY type complete
TITLE hypothetical protein Fl_orf879 - Mycoplasma pneumoniae (ATCC
ORGANISM 29342) (SGC3)
#formal_name Mycoplasma pneumoniae
ATCC 29342
#variety ATCC 29342
DATE 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
17-Jul-1998
ACCESSIONS S73757
REFERENCE S73757
#authors Himmelsreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li,
B.C.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:4420-4449
#title Complete sequence analysis of the genome of the bacterium
Mycoplasma pneumoniae.
#cross-references M0ID:97105885
#accession S73757
#status preliminary; nucleic acid sequence not shown;
translation not shown
```

#authors Paquin, B.; Lang, B.F.
#journal J. Mol. Biol. (1996) 255:688-701
#title The mitochondrial DNA of Allomyces macrognus: the complete genomic sequences from an ancestral fungus.

#cross-references MUID:96226032
#accession S63638
#status

#molecule_type DNA
#residues 1-382 #label PAQ
#cross-references EMBL:U01288; NID:gl236403; PIDN:AAC49221.1;
PID:gl236404
the nucleotide sequence was submitted to the EMBL Data Library, November 1995

#note

GENETICS
#gene
#genome
#introns

CLASSIFICATION
b6 homology; cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin--plastocyanin reductase 17K protein homology
electron transfer; heme; iron; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory chain

KEYWORDS
FEATURE
10-340 *domain cytochrome b homology #label CBH\
10-210 *domain cytochrome b6 homology #label CB6\
222-340 *domain plastocyaninol--plastocyanin reductase 17K protein homology #label 17K\
82,197 *binding_site heme iron, low potential (His) (axial ligands) #status Predicted\
96,202 *binding_site heme iron, high potential (His) (axial ligands) #status Predicted
#length 382 #molecular-weight 43467 #checksum 6973

SUMMARY
Query Match 16.7%; Score 93; DB 2; Length 382;
Best Local Similarity 33.3%; Pred. No. 9,25e-01;
Matches 17; Conservative 15; Mismatches 16; Indels 3; Gaps 3;

Ddb 8 PVLISANSLDPSLPNSHYIWNFGSLG-LCIVTQIVGTWLAHYAPS 57
|::| |::| |::| |::| |::| |::| |::| |::| |::|
YQ8142B.03; protein YDR211w; translation initiation factor eIF-2B homolog

Oy 22 PGMTLOLK-LAESFL-TNSSTESSFLETLERLCMLLEPSTGLTHARS 70
|::| |::| |::| |::| |::| |::| |::| |::| |::|

RESULT 3
ENTRY A48156 #type complete
TITLE translation regulator GCD6 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES guanine nucleotide exchange factor chain GCD6; protein YD8142B.03; protein YDR211w; translation initiation factor eIF-2B homolog
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 03-May-1994 #text_change 06-Feb-1998

ACCESSIONS A48156; S61578; S30776
REFERENCE A48156
Bushman, J.L.; Asuru, A.I.; Watts, R.L.; Hinnebusch, A.G. Mol. Cell. Biol. (1993) 13:1920-1932
Evidence that GCD6 and GCD7, translational regulators of GCNA, are subunits of the guanine nucleotide exchange factor for eIF-2 in Saccharomyces cerevisiae.

#cross-references MUID:93180841
#accession A48156
#molecule_type DNA
#residues 1-712 #label BUS
#cross-references EMBL:I07115; NID:g171572; PID:g171574
sequence extracted from NCBI backbone (NCBIN:126018, NCBP:126021)

#note S61576
Oliver, K.; Harris, D.
submitted to the EMBL Data Library, December 1995

#authors Olmer, K.; Harris, D.
#submission S61578
#accession S61578
#molecule_type DNA
#residues 1-712 #label OLI
#cross-references EMBL:268195; NID:g1122341; PID:e213795; PID:g1122344; MIPS:YDR211w

FEATURES	source
47023	gap of unknown length
51524:	contig of 4502 bp in length
51525	gap of unknown length
53246:	contig of 1722 bp in length
53247	gap of unknown length
56192:	contig of 2946 bp in length
56193	gap of unknown length
58896:	contig of 2704 bp in length
58897	gap of unknown length
61546:	contig of 2850 bp in length
61547	gap of unknown length
65573:	contig of 4127 bp in length.
Location/Qualifiers	
1. .65673	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="16"	
/clone="RPC1-11.529K1"	
16932 a 16632 c 16336 g 15686 t	87 others

Query Match	8.34;	Score 34.6;	DB 42;	Length 65673;
Best Local Similarity	54.34;	Prod. No. 7.2;		
Matches 70;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;
190	TGAAGAAGCTCTGCTCCTCTCCATCTCCCTTCAGGACACAGGCTCACCCCTCCACCATGC	249		
39815	TGANAATCTTCATCCACTCCGCTCCCTCTCTGGACGGGGACCCCTCGACCCCTTC	39756		
250	AAGATCTCAACACCATCTTGTCTGCAACACATCAGACGCATTGAAGCCTGTGTCCTCTT	309		
39755	TAACTTCAAAGCCCAATCTCTGTGATGCACGCGCTCTCTTTTGAAGAAATTATCCCTTC	39696		
310	GGCCCGGGC	318		
39695	GATCTGGC	39687		

RESULT 12	HSJD144C9/c	HSJD144C9	77322 bp	DNA	PRI	22-NOV-1999
LOCUS	HSJD144C9/c	Human DNA sequence from clone RPI-144C9 on chromosome 1p34.3-36.11,				
DEFINITION		complete sequence.				
ACCESSION		AL096774				
VERSION		AL096774.9	GI:6465842			
KEYWORDS		HTG.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
AUTHORS		1 (bases 1 to 77322)				
		Hall P				

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (22-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk

On Nov 22, 1999 this sequence version replaced gi:6138775.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi, EMBL; Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information

on the WormPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RPL-144C5 is from the library RPL-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTATOR: pCYPAC2 This sequence is the entire insert of clone RPL-144C9.

```

Location/Qualifiers
1. .77322
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="1"
   /map="p34.3-36.11"
   /clone_rpi="I44C9"
   /Clone_lib="RPC1-1"
   complement(1710..2227)
   /note="match: GSS: Em:AQ355665"
   complement(1780..2227)
   /note="match: GSS: Em:AQ355618"
   2858..3151
   /note="match: GSS: Em:AQ627110"
   complement(1143..1190)
   /note="match: GSS: Em:AQ627110"
   complement(2110..2100)
   /note="match: GSS: Em:AQ627110"
   complement(2161..2100)
   /note="match: GSS: Em:AQ627110"
   2101..2100
   /note="match: GSS: Em:AQ627110"
   complement(2101..2100)
   /note="match: GSS: Em:AQ627110"
   complement(2101..2100)
   /note="match: GSS: Em:AQ627110"
   complement(2101..2100)
   /note="match: GSS: Em:AQ627110"

```

```

Query Match      8 30 Score 34.4 DB 10 Length 77322
Best Local Similarity = 63.1% Pred. No. 8.4
Matches 53; Conservative 0 Mismatches 31 Index 957
QY    14 TTTCTCCATCCGAGCGCAGTGGCCACTATGGGTCTGGCCCCCTTCCTCT
Db    31300 TGTTGCCCTCCCTCCGAGGATCCCCTTTGGTGAGTATGGTGTTCAGGTACCAACCAC 31291

```

QY	74	CCTCTGACCTCCTCTGGAGCTCA	97
Db	31240	CTCTAGTAGTACCCTTCAGGCACACA	31217
RESULT	13	D86631	3626 bp DNA ROD
LOCUS	D86631/c	Mus sp.	DNA for oxytocin receptor, partial cds.
DEFINITION	D86631	Mus sp.	DNA for oxytocin receptor, partial cds.
ACCESSION	D86631	Mus sp.	DNA for oxytocin receptor, partial cds.
VERSION	D86631.1	GI:1902965	
KEYWORDS	oxytocin receptor		
SOURCE	Mus sp. (strain:129SV)	ES cell DNA.	
ORGANISM	Mus sp.		
	Eukaryota; Metazoa; Chordata; Vertebrata;		
	Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	Kubota,Y.		
AUTHORS	Kubota,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUL-1996)		
	to the DDBJ/EMBL/GenBank databases.		
	Yasue		

Kubota, Osaka University Medical School, Department of Obstetrics and Gynecology; 2-2 Yamadaoka, Suita, Osaka 565, Japan (E-mail: kubota@ygyne.med.osaka-u.ac.jp, Tel:81-6-879-3351, Fax:81-6-879-3359)

2 (bases 1 to 3626)

Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Saiji, F. and Murata, Y.

Structure and Expression of the Mouse Oxytocin Receptor Gene Unpublished (1996)

Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Saiji, F. and Murata, Y.

Structure and expression of the mouse oxytocin receptor gene Mol. Cell. Endocrinol. 124 (1-2), 25-32 (1996)

97179034 Location/Qualifiers

1..3626

/organism="Mus sp."

/strain="129SV"

/db_xref="taxon:10095"

/cell_type="ES cell"

1..1530

1531..1798

/number=1

1799..2240

/number=1

2241..2393

/number=2

2394..2485

/number=2

2486..3547

/number=3

2629..24546

/codon_start=1

/product="oxytocin receptor"

/protein_id="BAI8936.1"

/db_xref="GI:1902986"

/translation="MEGTPAANNSIEDLSCVPPGAGNLTAGPRNEALARVEVA VLGLIFLALSGMACVLLALTRHSHLSIADLYVAFTVQLPQALADIT FRFGPDLLCRVLYQVGVFASFTYLLMLSDRCIAICQPLSRNRTDRLAVAT WLGLVASVQVHIFSLFVAGVDFCNVFTOPWGPAYVITLAVITVPVYLLA CYGLISFKLNOLRLKTAAMAAAGSDAAGGAGGAALRVSSVLLKISKIRIVMT RIVLAFIVQWTFPPFFVQMSVDVNAPE"

3548..3626

/number=3

BASE COUNT 782 a 921 c 1131 g 792 t

ORIGIN

Query Match 8.2%; Score 34.2; DB 12; Length 3626;

Best Local Similarity 60.0%; Pred. No. 7.8;

Matches 57; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 152 ACAATTCCTCTATGAGTCAGCTTCCTGGATTCGTTGAAGCTCTGCTCTCTC 211

Db 2085 AAAAGCTTACTACGGGTACATCTCTGGGGTCCGCCAAGAGCCATCTTTCNC 2026

Qy 212 CATCTCCCTTCAGGACGAGGCTACCCCTCCCA 246

Db 2025 CTCCTGCTTCAGACTAGCCGAGCCAGCTCCCA 1991

RESULT 14

HSJ858B16 36676 bp DNA PRI 23-NOV-1999

LOCUS

DEFINITION

Human DNA sequence from clone 858B16 on chromosome 22. Contains the KIAA0542 gene and a gene for a novel protein similar to hamster PSSC (phosphatidylserine Decarboxylase Proenzyme, EC 4.1.1.65). Contains ESTs, GSSs and a putative CpG island, complete sequence.

AL096768

AL096768.7 GI:5596770

FTG: CpG Island; KIAA0542; Phosphatidylserine Decarboxylase Proenzyme; PSSC.

human

ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 36676)

Barlow, K.

Direct Submission

Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonequest@sanger.ac.uk

On Jul 27, 1999 this sequence version replaced gi:5579004.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi., EMBL; Sw., SWISSPROT; Tr., TRMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr22

858B16 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR: PCPAC2

IMPORTANT: This sequence is not the entire insert of clone 858B16. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 694E4 (AL031255) is at 36572 in this sequence.

Location/Qualifiers

1..36676

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22"

/clone_lib="RPCI-5"

/clone="RPS-858B16"

join(<273..454,1413..1488,1978..2073,2538..2614,3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,11023..11199,11271..11524,12717..12948,14414..14546,14669..14747,15784..15895,15983..16217)

/gene="dJ858B16.1"

/note="match: ESTs: Em:R53362 Em:278408 Em:AA521233 Em:R53964 Em:AI806913 Em:AA158770 Em:AA83285 Em:AA974403 Em:AA594535 Em:AI858196 Em:AI425026 Em:R51102 Em:AA52394 Em:AA529254 Em:AA553345 Em:AI809747 Em:AA846853 Em:AA971234 Em:R54651 Em:AI141022 Em:RA159572 Em:AI361548 Em:AA916775 Em:AI191287 Em:AV031353 Em:AA779744 Em:N64034 Em:AV075750 Em:AV095355 Em:AA865102 Em:AT232370 Em:AA595561 Em:AA088219 Em:W25871 Em:N56896 Em:W90666 Em:AV030271 Em:AV116402 Em:AV122327 Em:AI024249 Em:AA673371 Em:AI190893 Em:AI019759 Em:AI770152 Em:RA49582 Em:N30173 Em:AI342072 Em:R54847 Em:AI126154 Em:AA18158 Em:AA044074 Em:N54928 Em:AA215582 Em:AA398600 Em:AI623440 Em:AI024124 Em:RA637284 Em:AI824779 Em:AI094871 Em:AI192647 Em:N63934 Em:AV083921"

/evidence="not_experimental"

/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"

join(<273..454,1413..1488,1978..2073,2538..2614,3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,11023..11199,11271..12948,14414..14546,14669..14747,15784..15895,15983..16218)

mRNA

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2000, 14:42:21 ; Search time 479.47 Seconds
(without alignments)
-1139.903 Million cell updates/sec

Title: US-09-092-296-3
Perfect score: 180
Sequence: 1 CAGGAGCCAGTGGCCACTA.....CTTCAAAAGCTGCGCTCCT 180

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database: GenEmbl.*

Word size: 0

Number of hits that pass the threshold: 1642386

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pri.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sv.*
- 15: gb_un.*
- 16: gb_v1.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_v1.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: gb_ba1.*
- 37: gb_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*
- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*

50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	90.8	50.4	192650	10	AB023048	AB023048 Homo sapi	
2	90.8	50.4	200000	10	AF000511	AF000511 Homo sapi	
3	90.8	50.4	47323	11	AC005937	AC005937 Homo sapi	
4	33.4	18.6	97580	11	HS91N18	AL031594 Human DNA	
5	32.2	17.9	36676	11	HS785B16	AL096768 Human DNA	
C	31.2	17.3	177241	11	HS402G11	AL022328 Human DNA	
	31.2	17.3	303319	42	AC006927	AC006927 Homo sapi	
8	31	17.2	97339	44	AB011039	AB011039 Homo sapi	
9	30.8	17.1	5280	9	AB011114	AB011114 Homo sapi	
10	30.8	17.1	168986	40	AC008013	AC008013 Homo sapi	
11	30.8	17.1	110000	43	AC007623_0	AC007623 Homo sapi	
12	30.8	17.1	179622	43	AC009533	AC009533 Homo sapi	
13	30.4	16.9	1091	7	SAGAPDHR	X04301 Mustard mRN	
C	30.4	16.9	1301	12	AF059567	AF059567 Mus muscu	
	30.4	16.9	1098	12	MMCARRH	Y10320 M.musculus	
15	30.4	16.9	1669	12	MMCARRHOM	Y11929 M.musculus	
16	30.4	16.9	1131	12	MMINKR4B01	U66084 Mus musculu	
C	17	30.4	16.9	1206	12	MMU90715	U90715 Mus musculu
18	30.4	16.9	1206	12	MMU90715	U93354 Thermotoga	
C	19	30.2	16.8	2803	2	TNU93354	U93354 Thermotoga
20	30	16.7	751	7	ATRNABBC1	X75162 A.thaliana	
C	21	30	16.7	85109	8	ATT2J13	AL132967 Arabidops
22	30	16.7	4341	9	HUMHG2A	M83665 Human high	
C	23	30	16.7	124990	11	HS434P1	Z97056 Human DNA s
24	30	16.7	159125	44	AC015713	AC015713 Homo sapi	
25	29.8	16.6	42259	6	LBPHIGIE	X98106 Lactobacill	
26	29.8	16.6	6042	6	LBPHIGP49	X90510 Lactobacill	
C	27	29.8	16.6	2160	9	D84361	D84361 Human mRNa
C	28	29.8	16.6	154791	45	AC015478	AC015478 Homo sapi
29	29.6	16.4	1691	4	CHKC1PA2M	J00838 Chicken typ	
30	29.6	16.4	1691	4	GGCO10	V00390 Messenger f	
31	29.6	16.4	1683	4	AP000346	AP000346 Homo sapi	
C	32	29.6	16.4	151600	9	ESTALDNA	X81027 H.sapiens t
33	29.6	16.4	1000	9	HS21018	AL035405 Human DNA	
C	34	29.6	16.4	188016	33	AC007030	AC007030 Homo sapi
35	29.6	16.4	204617	35	AC006074	AC006074 Drosophil	
36	29.4	16.3	1755	10	HS297017	Z97017 Homo sapien	
C	37	29.4	16.3	149698	11	HS76C18	Z98755 Human DNA s
38	29.4	16.3	135545	12	MHHC213L3	AF109905 Mus muscu	
C	39	29.4	16.3	168609	41	AC009297	AC009297 Homo sapi
40	29.4	16.3	291288	42	AC008878	AC008878 Homo sapi	
41	29.4	16.3	106687	43	AC015892	AC015892 Mus muscu	
42	29.4	16.3	208079	44	AC013391	AC013391 Homo sapi	
43	29.4	16.3	259307	45	AC015657	AC015657 Mus muscu	
C	44	29.4	16.3	31311	45	AC017192	AC017192 Drosophil
C	45	29.2	16.2	94179	42	AC010028	AC010028 Drosophil

ALIGNMENTS

RESULT 1
AB023048
LOCUS AB023048 192650 bp DNA PRI 20-NOV-1999
DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
clone:53L9, complete sequence.
ACCESSION AB023048.1 GI:5672603
VERSION HTG.
KEYWORDS Homo sapiens
SOURCE Homo sapiens
ORGANISM Eukaryota; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS

1 (sites)
Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Matanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Bahram,S. and Inoko,H.
Molecular dynamics of MHC genes unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
20027539

MEDLINE
REFERENCE
AUTHORS

2 (bases 1 to 192650)
Shina,T. and Takishima,N.
Direct Submission
Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shina, Tokai University School of Medicine, Department of
Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail:tshina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)

FEATURES

Location/Qualifiers
1..192650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="978SR"
/chromosome="6"
/clone="5319"
/map="6p21.3"

BASE COUNT 49862 a 44743 c 45833 g 52212 t

ORIGIN

Query Match 50.4%; Score 90.8; DB 10; Length 192650;
Best Local Similarity 97.9%; Pred. No. 3.6e-19;
Matches 92; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 78 CAGGCGCGGTATGACCTTGCACCTGACCTGAGGAGCTTTCTGCAAGTCTCTCT 137
|||||
Db 177979 CAGGCGCGGTATGACCTTGCACCTGACCTGAGGAGCTTTCTGCAAGTCTCTCT 178038

QY 138 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGCT 171
|||||
Db 178039 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGT 178072

RESULT 2

AP000511 200000 bp DNA PRI 28-SEP-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
DEFINITION section 10/20.
ACCESSION AP000511
VERSION AP000511.1 GI:5926698
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (sites)
Shina,S., Tamiya,G., Oka,A. and Inoko,H.
Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
Published Only in Database (1999) in press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)

COMMENT

This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:81-463-93-1121, Fax:81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS

project of JST

Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www-alis.tokyo.jst.go.jp.

FEATURES
SOURCE

1..200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
108774..108864
/standard_name="D6S2088"
/note="SHGC-12985:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735268"

STS

115648..115906
/standard_name="D6S1898"
/note="WI-9418:The location is between each flanking site
of PCR primers."
/db_xref="GDB:678272"
/note="Cda0vhl0:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443382"
142030..142159
/standard_name="D6S2203"
/note="SHGC-16870:The location is between each flanking
site of PCR primers."
/db_xref="GDB:741215"

STS

142322
/note="RH18132:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4573021"
complement(142479..142740)
/standard_name="D6S1851"
/note="SHGC-10808:The location is between each flanking
site of PCR primers."
/db_xref="GDB:675281"

STS

complement(164525..164665)
/note="SHGC-3064:The location is between each flanking
site of PCR primers."
/db_xref="GDB:1234116"
complement(169209..169574)
/standard_name="D6S952"
/note="Jt5233:The location is between each flanking site
of PCR primers."
/db_xref="GDB:313481"

STS

BASE COUNT 52605 a 47531 c 49366 g 50498 t
ORIGIN

Query Match 50.4%; Score 90.8; DB 10; Length 200000;
Best Local Similarity 97.9%; Pred. No. 3.6e-19;
Matches 92; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 78 CAGGCGCGGTATGACCTTGCACCTGACCTGAGGAGCTTTCTGCAAGTCTCTCT 137
|||||

Db 110306 CAGGCGCGGTATGACCTTGCACCTGAGGAGCTTTCTGCAAGTCTCTCT 110365
|||||

QY 138 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGCT 171
|||||

Db 110366 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGT 110399
|||||

RESULT 3

AC005937
LOCUS AC005937 47323 bp DNA PRI 05-NOV-1998
DEFINITION Homo sapiens clone UMGc:370M23.002 from 6p21, complete sequence.
ACCESSION AC005937
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.


```

Matches 92; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 78 CAGGGCGGGTATGACTTTCACACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 137
|||||
Db 35465 CAGGGCGGGTATGACTTTCACACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 35524
|||||
QY 138 ATGAGTCCAGCTCTCTGAAATTCCTGAAAAGCT 171
|||||
Db 35525 ATGAGTCCAGCTCTCTGAAATTCCTGAAAAGGT 35558
|||||

RESULT 4
HS591N18
LOCUS
DEFINITION
Human DNA sequence from clone 591N18 on chromosome 22q13.1-13.2.
Contains a COX6B (Cytochrome C Oxidase subunit VIb (EC 1.9.3.1))
pseudogene, ESis, GSSs and two putative CpG islands, complete
sequence.
ACCESSION
AL031594.9 GI:5050980
KEYWORDS
HTG; COX6B; CpG Island; Cytochrome C Oxidase.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 97580)
Steward,C.
Direct Submission
Submitted (25-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 11, 1999 this sequence version replaced gi:4914529.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM: ENBL; SW: SWISSPROT; Tr: TREMBL
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
591N18 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: PCYPAC2
IMPORTANT: This sequence is not the entire insert of clone 591N18.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 229A8 (286090) is at 97475 in this
sequence. The true right end of clone 1042K10 (AL022238) is at
34197 in this sequence.
FEATURES
source
location/Qualifiers
1..97580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q13.1-13.2"
/clone="RP4-591N18"
/clone_lib="RPC1-4"
1..227
/Note="AluY repeat: matches 75..300 of consensus"
repeat_region
996..1104
/Note="FLAN_C repeat: matches 1..108 of consensus"
repeat_region
1147..1220
/Note="AluU/FRAM repeat: matches 217..294 of consensus"
repeat_region
1225..1501
/Note="AluJo repeat: matches 12..282 of consensus"
repeat_region
1552..1864
/Note="AluY repeat: matches 1..311 of consensus"
repeat_region
1869..2134
/Note="AluJb repeat: matches 21..275 of consensus"
repeat_region
2942..3234
/Note="AluSc repeat: matches 4..295 of consensus"
repeat_region
4854..5144
/Note="AluJb repeat: matches 1..308 of consensus"
misc_feature
6310..6584
/Note="match: GSS AQ038173"
repeat_region
6506..6575
/Note="MER3 repeat: matches 134..209 of consensus"
repeat_region
6585..6665
/Note="FLAN_A repeat: matches 27..116 of consensus"
repeat_region
6673..6694
/Note="MER3 repeat: matches 148..167 of consensus"
repeat_region
6695..6988
/Note="AluSx repeat: matches 1..290 of consensus"
repeat_region
6989..7068
/Note="MER3 repeat: matches 58..148 of consensus"
repeat_region
7497..7761
/Note="AluJb repeat: matches 1..301 of consensus"
repeat_region
7845..8136
/Note="AluSx repeat: matches 1..299 of consensus"
repeat_region
8467..8776
/Note="AluY repeat: matches 1..311 of consensus"
repeat_region
9209..9505
/Note="AluSx repeat: matches 1..295 of consensus"
repeat_region
11780..12007
/Note="L1ME1 repeat: matches 5465..5691 of consensus"
repeat_region
12018..12387
/Note="HERV23 repeat: matches 1184..1573 of consensus"
repeat_region
12454..12544
/Note="L1ME1 repeat: matches 5679..5763 of consensus"
repeat_region
12545..12843
/Note="AluJb repeat: matches 1..297 of consensus"
repeat_region
12844..13179
/Note="L1ME1 repeat: matches 5763..6159 of consensus"
repeat_region
13240..13373
/Note="AluJo repeat: matches 1..136 of consensus"
repeat_region
13408..13612
/Note="AluSg/x repeat: matches 96..293 of consensus"
repeat_region
13625..13924
/Note="AluSx repeat: matches 1..301 of consensus"
repeat_region
13935..14230
/Note="AluSx repeat: matches 1..297 of consensus"
repeat_region
14237..14466
/Note="MIR repeat: matches 11..261 of consensus"
repeat_region
14729..14785
/Note="MIR repeat: matches 202..261 of consensus"
repeat_region
15147..15467
/Note="L2 repeat: matches 2370..2690 of consensus"
repeat_region
15625..15910
/Note="AluSg repeat: matches 1..296 of consensus"
repeat_region
16000..16300
/Note="AluSg repeat: matches 1..291 of consensus"
repeat_region
16422..16732
/Note="AluJb repeat: matches 1..311 of consensus"
repeat_region
16733..16813
/Note="L2 repeat: matches 2667..2746 of consensus"
misc_feature
complement(16814..17293)
/Note="match: GSSs AQ377982 AQ380087"
repeat_region
16907..17100
/Note="MIR repeat: matches 2..197 of consensus"
repeat_region
18063..18376
/Note="AluJb repeat: matches 1..307 of consensus"
repeat_region
18462..18537
/Note="738 copies 2 mer tt 78% conserved"
repeat_region
18540..19522

```

misc_feature /note="SVA repeat: matches 2. .954 of consensus" 19061. .20976
 /note="Random repeat. Some base discrepancies edited. Assembly is consistent with Restriction Digest." 19115. .20959
 misc_feature /note="CpG island" /evidence=not_experimental 19540. .20001
 repeat_region /note="SVA repeat: matches 521. .954 of consensus" 20019. .20360
 repeat_region /note="SVA repeat: matches 521. .845 of consensus" 20361. .20854
 repeat_region /note="SVA repeat: matches 519. .994 of consensus" complement(20855. .20945)
 misc_feature /note="match: GSS AQ473058" 21493. .21658
 repeat_region /note="AluX repeat: matches 1. .165 of consensus" 21793. .22086
 repeat_region /note="AluX repeat: matches 1. .297 of consensus" 22379. .22534
 repeat_region /note="FRAM repeat: matches 1. .162 of consensus" 23688. .24049
 repeat_region /note="MLT1A1 repeat: matches 1. .365 of consensus" 24100. .24408
 repeat_region /note="AluYb8 repeat: matches 1. .305 of consensus" 25227. .25536
 repeat_region /note="AluYb repeat: matches 1. .310 of consensus" 25556. .25779
 repeat_region /note="AluYb repeat: matches 58. .297 of consensus" 25834. .26147
 repeat_region /note="AluSg repeat: matches 1. .307 of consensus" 26148. .26192
 repeat_region /note="MIR repeat: matches 187. .230 of consensus" 26377. .26688
 repeat_region /note="AluX repeat: matches 2. .311 of consensus" 26884. .27010
 repeat_region /note="FLAM_C repeat: matches 1. .133 of consensus" 30237. .30344
 repeat_region /note="MIR repeat: matches 103. .206 of consensus" 31467. .31777
 repeat_region /note="AluSg repeat: matches 1. .310 of consensus" 31963. .32080
 repeat_region /note="AluYb repeat: matches 1. .118 of consensus" 32118. .32408
 repeat_region /note="AluYb repeat: matches 1. .292 of consensus" 32623. .32927
 repeat_region /note="AluYb repeat: matches 1. .300 of consensus" 33031. .33354
 repeat_region /note="AluSg repeat: matches 1. .304 of consensus" 33381. .33519
 repeat_region /note="L2 repeat: matches 2572. .2709 of consensus" 33560. .33777
 repeat_region /note="AluX repeat: matches 1. .219 of consensus" 34284. .34597
 repeat_region /note="AluX repeat: matches 1. .312 of consensus" 35062. .35370
 repeat_region /note="AluJ repeat: matches 1. .309 of consensus" 35600. .35753
 repeat_region /note="MIR repeat: matches 2. .230 of consensus" 35782. .35854
 repeat_region /note="MER21B repeat: matches 720. .790 of consensus" 35912. .36626
 repeat_region /note="MER21B repeat: matches 1. .713 of consensus" 36860. .36969
 repeat_region /note="Alu repeat: matches 1. .308 of consensus" 37152. .37273
 repeat_region /note="AluSg/x repeat: matches 185. .302 of consensus"

Query Match 18.69; Score 33.4; DB 11; Length 97580;
 Best Local Similarity 57.08; Pred No. 1.6; Mismatches 46; Indels 0; Gaps 0;
 Matches 61; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
 60.TTGACGCTCACATGACACAGCGGGTATGATCTTTCGACTGTAAGCTGAAGGATCTT 119

Db 2606 TTCCCTCCATCCCTCCAGGTTTAGTAAAGACTTTTCAAAAGCAAGGTAGGTGTAAC 2665
 QY 120 TTCTGACAGAGTTCCTCTATGAGTCCAGCTCCCTGGAAATGCTTGA 166
 Db 2666 TGTAGAAGATTCATGATATAGTACCGCACTTTGTATTCTCTGAA 2712

RESULT 5
 HSJ858B16 36676 bp DNA PRI 23-NOV-1999
 LOCUS
 DEFINITION Human DNA sequence from clone 858B16 on chromosome 22. Contains the KIAA0542 gene and a gene for a novel protein similar to hamster PSSC (Phosphatidylserine Decarboxylase Proenzyme, EC 4.1.1.65). Contains ESTs, GSSs and a putative CpG island, complete sequence.
 ACCESSION AL096768.7 GI:5596770
 VERSION
 KEYWORDS HTG; CpG Island; KIAA0542; Phosphatidylserine Decarboxylase Proenzyme; PSSC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 36676)
 AUTHORS Barlow,K.
 TITLE Direct Submission
 JOURNAL Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequests@sanger.ac.uk
 COMMENT On Jul 27, 1999 this sequence version replaced gi:5579004. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22
 858B16 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://pacpac.med.buffalo.edu/VECTOR:pcypac2 IMPORTANT: This sequence is not the entire insert of clone 858B16. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 The true left end of clone 694E4 (AL031255) is at 36572 in this sequence.
 FEATURES
 source
 1. .36676
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone_id="RPCI-5"
 /clone="RP5-858B16"
 join(273, 454, 1413, 1488, 1978, 2073, 2538, 2614, 3997, 4099, 5606, 5702, 8812, 8972, 9435, 9509, 10811, 10885, 11023, 11199, 11271, 11524, 12717, 12948, 14414, 14546, 14669, 14747, 15784, 15895, 15983, 16217)
 /gene="dJ858B16.1"
 /note="match: ESTs: Em:R53362 Em:278408 Em:AA521233...
 ...
 mRNA


```

repeat_region 24501..24747
/note="L1MEC repeat: matches 1501..1751 of consensus"
repeat_region 24978..25194
/note="L1MEC repeat: matches 272..492 of consensus"
repeat_region 25933..26246
/note="AluY repeat: matches 1..309 of consensus"
repeat_region 27926..28225
/note="AluX repeat: matches 2..302 of consensus"
repeat_region 28694..28997
/note="AluSg repeat: matches 1..304 of consensus"
repeat_region 29001..29119
/note="AluJo/FLAM repeat: matches 3..133 of consensus"
repeat_region 29859..30170
/note="L1ME repeat: matches 5189..5489 of consensus"
repeat_region 30171..30449
/note="AluX repeat: matches 24..304 of consensus"
repeat_region 30450..30556
/note="L1ME repeat: matches 5086..5189 of consensus"
repeat_region 30567..30813
/note="L2 repeat: matches 1079..1303 of consensus"
repeat_region 30975..31047
/note="L1ME/D repeat: matches 5325..5397 of consensus"
repeat_region 31048..31362
/note="AluY repeat: matches 1..303 of consensus"
repeat_region 31363..31990
/note="L1ME/D repeat: matches 4762..5325 of consensus"
repeat_region 31997..32351
/note="MLT/Al repeat: matches 16..365 of consensus"
repeat_region 32365..32907
/note="L1ME/D repeat: matches 4184..4755 of consensus"
repeat_region 32908..33208
/note="AluX repeat: matches 1..301 of consensus"
repeat_region 33209..33464
/note="L1ME/D repeat: matches 3936..4184 of consensus"
repeat_region 33468..33637
/note="FRAM repeat: matches 1..175 of consensus"
repeat_region 33650..33707
/note="L1MEC repeat: matches 2384..2075 of consensus"
repeat_region 33708..33998
/note="AluX repeat: matches 1..288 of consensus"
repeat_region 33999..34144
/note="L1MEC repeat: matches 2231..2385 of consensus"
repeat_region 34781..34916
/note="4 copies 34 mer 87% conserved"
repeat_region 34792..34910
/note="7 copies 17 mer 74% conserved"
repeat_region 35137..35466
/note="AluSg repeat: matches 1..310 of consensus"
repeat_region 35626..35922
/note="AluSg repeat: matches 1..297 of consensus"
repeat_region 35932..36042
/note="L2 repeat: matches 2643..2750 of consensus"
repeat_region 36532..37622
/note="L1MEC repeat: matches 255..1397 of consensus"
repeat_region 37623..37663
/note="AluSg repeat: matches 208..248 of consensus"
repeat_region 37706..37865
/note="AluSg/X repeat: matches 127..286 of consensus"
repeat_region 37881..38081
/note="L1ME5 repeat: matches 7541..7757 of consensus"
repeat_region 38099..38311

```

```

Query Match 17.3% Score 31.2; DB 11; Length 177241;
Best Local Similarity 57.0%; Pred. No. 8.7;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 16 CACTATGGGTCTGGGCTGCCCTTCCTCTTGACCTCTTGCGAGTCACATGG 75
|||||
Db 109678 CACAGGTCGCCAGCGGCTCTCGCGGCTCTCGAGCCATGCCACCCACATGG 109619
|||||

QY 76 AACAGGCGGGTATGACTTTGCACTGAAGCTGAAGAG 115
|||||
* Db 109618 AACATGCTGGGTGGACTGACGGATGGAGCTGGGCGAG 109579

```

```

RESULT 7
AC006927
LOCUS AC006927
DEFINITION Homo sapiens chromosome 12p13.1-17.1-21.3 clone RPC11-44J21, ***
SEQUENCING IN PROGRESS ***, 139 unordered pieces.
AC006927.19 GI:6091655
HTG; HTG_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 303319)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodda,B., Bouck,D., Bowie,S., Brooks,A., Buhey,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Desharo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogue,M.,
Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondajewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,K.,
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R., Tabor,P.,
Taylor,T., Vasquez,L., Vinson,R., Vo.O., Wabba,M., Watlington,S.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,K.
Direct Submission
Unpublished
TITLE Direct Submission
JOURNAL Direct Submission
AUTHORS Worley,K.C.
REFERENCE 2 (bases 1 to 303319)
JOURNAL Submitted (26-FEB-1999) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030 USA
On Oct 21, 1999 this sequence version replaced gi:5882346.
* NOTE: this is a 'working draft' sequence. It currently
* consists of 139 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 830: contig of 830 bp in length
* 831 1660: contig of 830 bp in length
* 1661 2498: contig of 838 bp in length
* 2499 3178: contig of 680 bp in length
* 3179 3829: contig of 651 bp in length
* 3830 4616: contig of 787 bp in length
* 4617 5463: contig of 847 bp in length
* 5464 6330: contig of 867 bp in length
* 6331 7341: contig of 1011 bp in length
* 7342 7832: contig of 491 bp in length
* 7833 8684: contig of 852 bp in length
* 8685 9411: contig of 727 bp in length
* 9412 10220: contig of 809 bp in length
* 10221 11025: contig of 805 bp in length
* 11026 11817: contig of 792 bp in length
* 11818 12571: contig of 754 bp in length
* 12572 13368: contig of 797 bp in length
* 13369 14228: contig of 860 bp in length
* 14229 15013: contig of 785 bp in length
* 15014 15730: contig of 717 bp in length
* 15731 16617: contig of 887 bp in length
* 16618 17335: contig of 718 bp in length
* 17336 17992: contig of 657 bp in length
* 17993 18287: contig of 295 bp in length

```


KEYWORDS
SOURCE
ORGANISM

HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168986)

REFERENCE
AUTHORS

Kuzny,D., Aronson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
Kondajewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
Logan,O., Lu,J., Marandel,I., Martinez,C., Merscher,S., Miller,A.,
Montgomery,K., Osvald,G., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
Vo,O., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W.,
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.

TITLE

Direct Submission

Unpublished

2 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (10-JUL-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (28-AUG-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (02-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (09-OCT-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Aug 28, 1999 this sequence version replaced gi:5739545.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

COMMENT

standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Contig length: 168986
Phrap values in estimate:
Average error rate (BCM-Phrap estimate): 0.000156145
Fraction of Phrap values less than 40 : 0.03878
Number of consensus changing edits: 14
Number of N's in consensus : 0

----- Consensus changing edits -----

Position	Original*Context	Edited*Context
4287	cctaaaccc(n)cagctctcca	cctaaaccc(n)cagctctcca
5004	cggctctca(n)tcgcgcctc	cggctctca(n)tcgcgcctc
5018	ccgctctgca(n)gggggcctc	ccgctctgca(n)gggggcctc
5050	atggggggccc(n)aagagcctg	atggggggccc(n)aagagcctg
86008	catagccca(n)ccttgctgt	catagccca(n)ccttgctgt
102306	gaccatcgtg(n)aagcccccag	gaccatcgtg(n)aagcccccag
102546	gggtaacccc(n)cttgccctt	gggtaacccc(n)cttgccctt
102547	ggtaacccc(n)cttgccctt	ggtaacccc(n)cttgccctt
109591	tttgttttg(n)tttgtttt	tttgttttg(n)tttgtttt
109596	tttgttttg(n)tttgtttt	tttgttttg(n)tttgtttt
155478	tgcatgtgtg(n)gtgcatatgc	tgcatgtgtg(n)gtgcatatgc
165391	agggaaagaa(n)aaggaagga	agggaaagaa(n)aaggaagga
165392	aggaaggaan(n)aggaagga	aggaaggaan(n)aggaagga
165993	aactcaaaa(n)gggtgcagca	aactcaaaa(n)gggtgcagca

----- Distribution of Quality < 40 Bases -----

# bases	5	10	15	20	25	30	35	40
1000								
900								
800								
700								
600								
500								
400								
300								
200								
100								
0								

Version: 1.01 xfgc.

Location/Qualifiers

1..168986
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12p12-21.3-21.8"
/clone="RPC111-55114"
507..655
/rpt_family="MIR"
complement(725..1166)
/rpt_family="MER110A"
981..1329
/note="Region: ol46h06.sl Homo sapiens cDNA, AA017071"
complement(1338..1761)
/rpt_family="MTRC"
2100..2251
/rpt_family="MER112"
3343..3431

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality


```

* 194817 194873: gap of unknown length
* 194874 196660: contig of 1787 bp in length
* 196661 196717: gap of unknown length
* 196718 198486: contig of 1769 bp in length
* 198487 198543: gap of unknown length
* 198544 200297: contig of 1754 bp in length
* 200298 200354: gap of unknown length
* 200355 202066: contig of 1712 bp in length
* 202067 202123: gap of unknown length
* 202124 203824: contig of 1701 bp in length
* 203825 203881: gap of unknown length
* 203882 205545: contig of 1664 bp in length
* 205546 205602: gap of unknown length
* 205603 207253: contig of 1651 bp in length
* 207254 207310: gap of unknown length
* 207311 208900: contig of 1590 bp in length
* 208901 208957: gap of unknown length
* 208958 210510: contig of 1533 bp in length
* 210511 210567: gap of unknown length
* 210568 212117: contig of 1550 bp in length
* 212118 212174: gap of unknown length
* 212175 213721: contig of 1547 bp in length
* 213722 213778: gap of unknown length
* 213779 215323: contig of 1545 bp in length
* 215324 215380: gap of unknown length
* 215381 216948: gap of unknown length
* 216949 218452: contig of 1504 bp in length
* 218453 218509: gap of unknown length
* 218510 219991: contig of 1482 bp in length
* 219992 220048: gap of unknown length
* 220049 221520: contig of 1472 bp in length
* 221521 221577: gap of unknown length
* 221578 223033: contig of 1456 bp in length
* 223034 223090: gap of unknown length
* 223091 224519: contig of 1429 bp in length
* 224520 224576: gap of unknown length
* 224577 225977: contig of 1401 bp in length
* 225978 226034: gap of unknown length
* 226035 227416: contig of 1382 bp in length
* 227417 227473: gap of unknown length
* 227474 228834: contig of 1361 bp in length
* 228835 228891: gap of unknown length
* 228892 230288: contig of 1337 bp in length
* 230289 230285: gap of unknown length
* 230290 231603: contig of 1318 bp in length
* 231604 231660: gap of unknown length
* 231661 232977: contig of 1317 bp in length
* 232978 233034: gap of unknown length
* 233035 234335: contig of 1301 bp in length
* 234336 234392: gap of unknown length
* 234393 235688: contig of 1296 bp in length
* 235689 235745: gap of unknown length
* 235746 236860: contig of 1115 bp in length
* 236861 236917: gap of unknown length
* 236918 238201: contig of 1284 bp in length
* 238202 238258: gap of unknown length
* 238259 239538: contig of 1280 bp in length
* 239539 239595: gap of unknown length
* 239596 240860: contig of 1265 bp in length
* 240861 240917: gap of unknown length
* 240918 242182: contig of 1265 bp in length
* 242183 242239: gap of unknown length
* 242240 243497: contig of 1258 bp in length
* 243498 243554: gap of unknown length
* 243555 244805: contig of 1251 bp in length
* 244806 244862: gap of unknown length
* 244863 246101: contig of 1239 bp in length
* 246102 246158: gap of unknown length
* 246159 247381: contig of 1223 bp in length
* 247382 247438: gap of unknown length
* 247439 248638: contig of 1200 bp in length
* 248639 248695: gap of unknown length

```

```

* 248696 249893: contig of 1198 bp in length
* 249894 249950: gap of unknown length
* 249951 251148: contig of 1198 bp in length
* 251149 251205: gap of unknown length
* 251206 252402: contig of 1197 bp in length
* 252403 252459: gap of unknown length
* 252460 252655: contig of 1196 bp in length
* 252656 253712: gap of unknown length
* 253713 254901: contig of 1189 bp in length
* 254902 254958: gap of unknown length
* 254959 256146: contig of 1188 bp in length
* 256147 256203: gap of unknown length
* 256204 256954: contig of 751 bp in length
* 256955 257011: gap of unknown length
* 257012 258197: contig of 1186 bp in length
* 258198 258254: gap of unknown length
* 258255 259436: contig of 1182 bp in length
* 259437 259493: gap of unknown length
* 259494 260675: contig of 1182 bp in length
* 260676 260732: gap of unknown length
* 260733 261911: contig of 1179 bp in length
* 261912 261968: gap of unknown length
* 261969 263137: contig of 1169 bp in length
* 263138 263194: gap of unknown length
* 263195 264350: contig of 1156 bp in length
* 264351 264407: gap of unknown length
* 264408 265560: contig of 1153 bp in length
* 265561 265617: gap of unknown length
* 265618 266751: contig of 1134 bp in length
* 266752 266808: gap of unknown length
* 266809 267938: contig of 1130 bp in length
* 267939 267995: gap of unknown length
* 267996 268117: contig of 1122 bp in length
* 268118 268174: gap of unknown length
* 268175 270289: contig of 1115 bp in length
* 270290 270346: gap of unknown length
* 270347 271458: contig of 1112 bp in length
* 271459 271515: gap of unknown length
* 271516 272627: contig of 1112 bp in length
* 272628 272684: gap of unknown length
* 272685 273794: contig of 1110 bp in length
* 273795 273851: gap of unknown length
* 273852 274952: contig of 1101 bp in length
* 274953 275009: gap of unknown length
* 275010 276109: contig of 1100 bp in length
* 276110 276166: gap of unknown length

Query Match      17.18; Score 30.8; DB 43; Length 110000;
Best Local Similarity 63.58; Pred. No. 12;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 2 AGGAGCGCAGTGGCCACTATGGGCTGTGGGCTGCCCCCTGTCTCTCTGTGACCTCTT 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75932 AGGCGCTGAGGACCCACTGTGCAGTCTGGGCTGCCCATCTGTGCCATCTCC 75991
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 GCGAGCTCACATGG 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75992 AGGGCTCTCAGGG 76005
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AC009533/c
LOCUS
DEFINITION Homo sapiens clone RP11-22B23, *** SEQUENCING IN PROGRESS ***, 26
SOURCE human.
ACCESSION AC009533
VERSION AC009533.6 GI:6143848
KEYWORDS HTG; HTGS_PHASE1.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179622)

```


Qy	8	GCAGTGGCCACATATGGGGTCTGGGCTGCCCTCTGCTCTCTCTGACCTCTCTGGCAGC	67
Db	182	GGGTTGGCCAGGCCGGGCTACTGCTGGCGCCCAACATGCCCTTGTGCCCGCTGTGGC	133
Qy	68	TCACATGGAACAGCGCGGGTATGACTTTTGCAACTG	103
Db	122	AGAAATGGTCTTTCGCGCGCGTGAAGTTGCTACAG	87
RESULT	15		
MMCARH		1098 bp	MRNA
LOCUS			
DEFINITION		M.musculus mRNA for coxsackievirus and adenovirus receptor (MmCARE)	11-MAR-1997
ACCESSION		Y10320	
VERSION		Y10320.1	
KEYWORDS		46 kDa receptor protein; coxsackie and adenovirus receptor	
SOURCE		house mouse	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		1 (bases 1 to 1098)	
TITLE		Bergelson,J.M.	
JOURNAL		Direct Submission	
REFERENCE		Submitted (03-JAN-1997) J.M. Bergelson, Dana-Farber Cancer Institute, Division of Infectious Diseases, 44 Binney Street, Boston, MA 02115, USA	
AUTHORS		2 (bases 1 to 1098)	
TITLE		Bergelson,J.M., Cunningham,J.A., Droguett,G., Kurt-Jones,E.A., Krithivas,A., Hong,J.S., Horwitz,M.S., Crowell,R.L. and Finberg,R.W.	
JOURNAL		Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5	
MEDLINE		Science 275 (5304), 1320-1323 (1997)	
FEATURES		97190109	
source		Location/Qualifiers	
gene		1..1098	
CDS		1..1098	
		/organism="Mus musculus"	
		/strain="C57Bl6J"	
		/db_xref="taxon:10090"	
		/tissue_type="liver"	
		/clone_lib="GibcoBRL (Cat 10656-015)"	
		1..1098	
		/gene="mCAR"	
		1..1098	
		/gene="mCAR"	
		/codon_start=1	
		/product="coxsackie and adenovirus receptor homologue"	
		/protein_id="CAA71368.1"	
		/db_xref="GI:1881467"	
		/db_xref="SPTREMBL:P97792"	
		/translation="MARLLCGVADTFITSGLSITTPTEORIEKAGETAYLVLSPEOGQDLIEWLSFDNVLQVIVLYSGDKIVDNYLDLKGVRHFTSNDIGASINYNLQISDIGNYQCKVKKAGVANKFLLTVLVPSCGTCFVDGSEETIGKPEKESDLPQEFQKLSDSQTHPTPLAEMTSPVSVKMASSEKSGYISCTIGSDQCHLNDVPPNSNAGTAVIGTLLALVGLFCHHKRAREKRYDKIFEDVPPKRSITAKSTISGNHSLGSMSPNMEGYSKTOYNQVPSDEFERAPAKRVAAPLMSBAGVPMVPAQSKDSIV"	
BASE COUNT	294 a	275 c	247 t
DRIGIN			
Query Match	16.9%	Score 30.4	DB 12; Length 1098;
Best Local Similarity	61.2%	Pred. No. 13;	
Matches	49;	CDservative 0;	Mismatches 31; Indels 0; Gaps
Qy	2	AGGAGCGCATGGCCACTATGGGCTCTGGGCTGCCCTCTGCTCTCTTACACCTCTT	61
Db	712	AGCATCTCGGGCGCGCTCATGAGAGCTGCTGGCCCTTGTGCTCATCGGGGCATCTC	771
Qy	.62	GCGAGCTCACATGGAAACAGG	81

Db 772 TTCGCTGTCACAGGAACG 791
| | | | | | | | | |

Search completed: March 21, 2000, 14:44:47
Job time: 3902 sec


```

RESULT 2
ID V28846 standard; DNA; 1128 BP.
AC V28846;
DE Mouse coxsackievirus and adenovirus receptor encoding DNA.
KW Mouse coxsackievirus; adenovirus; receptor; CAR; cardiac infection;
KW myocarditis; pericarditis; dilated cardiomyopathy; meningoencephalitis;
KW pancreatic infection; acute pancreatitis; gastrointestinal tract;
KW diabetes mellitus; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 1..1128
FT FT /*tag= a
FT FT /product= "coxsackievirus and adenovirus receptor"
FT FT /transl_except= (pos:1096..1098,aa:aaa)
FT FT /note= "Xaa= a stop codon, the sequence is shown as
FT FT continuing but gets unclear"
FN W09811221-A2.
PD 19-MAR-1998.
PF 12-SEP-1997; U16189.
PR 13-SEP-1996; US-026100.
PA (DAND ) DANA FARBER CANCER INST INC.
PI Bergelson JM, Finberg RW, Horowitz MS;
DR WPI: 98-207384/18.
DR P-PSDB; W57213.
PT DNA encoding coxsackie virus and adenovirus receptor - useful for
PT modulation of e.g. cardiac, pancreatic or gastrointestinal infection
PT Disclosure; Fig 14; 104pp; English.
PS The present sequence encodes mouse coxsackievirus and adenovirus
CC receptor (CAR). The present invention also describes: (1) a method for
CC modulating CAR expression comprising contacting the cell with an agent
CC which modulates CAR protein activity or CAR nucleic acid expression,
CC such that a cell associated activity is altered relative to a cell
CC associated activity of the cell in the absence of the agent; and (2) a
CC method for detecting the presence of CAR in a biological sample
CC comprising contacting a biological sample with an agent capable of
CC detecting CAR protein or mRNA such that the presence of CAR is detected.
CC Modulation of CAR is useful for treatment of cardiac infection, e.g.
CC myocarditis, pericarditis or dilated cardiomyopathy, or infection of the
CC central nervous system, e.g. a non-specific febrile illness or
CC meningoencephalitis, or pancreatic infection, e.g. acute pancreatitis,
CC or infection of the respiratory or gastrointestinal tract or childhood
CC onset diabetes mellitus. Probes derived from CAR nucleic acids are
CC useful for hybridisation assays, and antibodies raised against CAR
CC protein are useful for blocking CAR expression. Cell-free assays which
CC include combining CAR protein and a candidate/test compound are useful
CC in screening for drugs which interact with CAR protein.
SQ Sequence 1128 BP; 297 A; 286 C; 288 G; 257 T;

Query Match 17.8%; Score 32; DB 1; Length 1128;
Best Local Similarity 62.5%; Pred. No. 0.21;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 2 AGGACGCGAGTGGCGCTATGGGCTCTGGGCTGCCCTTGTCTCTCTTGACCCCTCTT 61
DB 712 ACGATCGCGGCGCGCTCATAGGAGCGTGTGGCTTGTGCTCATCGGGCCATCTC 771
QY 62 GGCACCTCATGTGACAGG 81
DB 772 TTCTGCTCTCAGAGAGG 791

RESULT 3
ID T02964/C
AC T02964 standard; cDNA; 580 BP.
DE Cell-cycle regulatory protein p15
DE Cell-cycle regulatory protein p15 cDNA.
KW Cell-cycle regulatory protein p15; cyclin-dependent kinase inhibitor;
KW CCR; gene therapy; transgenic animal; cancer; cell proliferation;
KW ss; ds.

```

```

OS Mus sp.
FH Key Location/Qualifiers
FT CDS 91..483
FT FT /*tag= a
FT FT W09528483-A1.
PD 26-OCT-1995.
PF 14-APR-1995; U04636.
PR 14-APR-1994; US-227371.
PR 25-MAY-1994; US-248812.
PR 14-SEP-1994; US-306511.
PR 29-NOV-1994; US-346147.
PA (COLD-) COLD SPRING HARBOR LAB.
PI Beach DH, Demetrick DJ, Hannon GJ, Serrano M;
DR WPI: 95-373798/48.
DR P-PSDB; R85118.
PT New cell cycle regulating proteins bind to cyclin dependent kinase -
PT and related nucleic acids, antibodies etc., used in diagnosis and
PT therapy of abnormal cell proliferation, degeneration etc.
PS Claim 43; Page 81-82; 109pp; English.
CC cDNA (T02964) coding for the mouse cell-cycle regulatory (CCR)
CC protein p15 (R85118) was isolated from an embryonal carcinoma
CC library using a probe based on a mouse CCR p13.5 cDNA (T02965).
CC The isolated cDNA can be used: to detect mutations in CCR
CC genes that lead to cell proliferation; to breed transgenic
CC animals to study cellular disorders involving CCR allele
CC mutation/misexpression; and to correct CCR-deficient cells
CC (gene therapy).
SQ Sequence 580 BP; 109 A; 187 C; 197 G; 87 T;

Query Match 16.9%; Score 30.4; DB 1; Length 580;
Best Local Similarity 57.3%; Pred. No. 0.56;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 8 GCATGTGCCATCATGGGTCTGGGCTGCCCTTGTCTCTTGACCTCTTGCGCAGC 67
DB 128 CGGTGTGCCAGCGCGCTGCTGCTGCTGCCGACATGCTTCCCTGCTGTGC 69
QY 68 TCACATGCAGACAGGCGGGGTATGACTTTGCAACTG 103
DB 68 AGAATGTGCTTCCGCGCGCGGTGAGATGCTTACAG 33

RESULT 4
ID V50430 standard; cDNA; 1301 BP.
AC V50430;
DT 07-DEC-1998 (first entry)
DE Mouse coxsackievirus and Ad2 and Ad5 receptor (MCAR) cDNA.
KW MCAR; coxsackievirus receptor; CVB; adenovirus; Ad2 receptor;
KW Ad5 receptor; mouse; infection; vaccine; gene therapy; ds.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 157..1215
FT FT /*tag= a
FT FT W09833819-A1.
PD 06-AUG-1998.
PF 30-JAN-1998; U01724.
PR 30-JAN-1997; US-036986.
PA (UTNY ) UNIV NEW YORK STATE.
PI Philipson L, Tomko RP;
DR WPI: 98-437397/37.
DR P-PSDB; W69698.
PT DNA encoding human receptor for adenovirus C and coxsackievirus B -
PT for preventing and treating viral infection and rendering cells
PT susceptible to transformation by adenoviral vectors in gene therapy
PS Claim 2; Page 68-70; 88pp; English.
CC This cDNA molecule codes for mouse MCAR protein (see W69698) that
CC serves as a cellular receptor for adenoviruses of the serotypes 2
CC and 5 (subgroup C) and for the group B coxsackieviruses (CVB). It
CC was obtained by screening a lambda phage expression cDNA library
CC with antiserum containing antibodies specific for mouse CVB
CC binding protein p46, and was used to identify human HCAR cDNA

```

CC (see V50429). The invention also provides host cells transformed
 CC with DNA molecules encoding HCAR or MCAR and methods of producing
 CC the recombinant proteins or their derivatives. These proteins,
 CC their extracellular domains, as well as oligopeptides (see
 CC W69699-708) which bind virus, are also provided. Isolated HCAR or
 CC MCAR proteins or their fragments or variants are used to prevent
 CC or treat virus infections. Expressing the DNA in cells which lack
 CC these viral receptors renders the cells susceptible to
 CC transformation by adenoviral vectors carrying genes for gene
 CC therapy.
 SQ Sequence 1301 BP; 355 A; 327 C; 322 G; 297 T;
 Query Match 16.9%; Score 30.4; DB 1; Length 1301;
 Best Local Similarity 61.2%; Pred. No. 0.76;
 Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 2 AGGAGCGCACTGGCCTATGCGGTCTGGGCTGCCCTTTCCTTCCTTCCTTCCTT 61
 Db 868 AGATCGCGGGCGCGCTCATAGGACGCTGCTGGCCCTTGTGCTATCGGGCCATCCTC 927
 QY 62 GCGAGCTCATATGGACAGG 81
 Db 928 TTCTGCTCATACAGGAACG 947
 RESULT 5
 V86365/c
 ID V86365 standard; cDNA; 291 BP.
 AC V86365;
 DT 27-APR-1999 (first entry)
 DE EST clone AR34.
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS Homo sapiens.
 SN W09845435-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06954.
 FR 10-APR-1997; US-835913.
 FA (GEM) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI: 99-070076/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 210; 63pp; English.
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.
 SQ Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;
 Query Match 16.6%; Score 29.8; DB 1; Length 291;
 Best Local Similarity 70.2%; Pred. No. 0.69;
 Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 88 TATGACTTTCAGCTGAAGCTGAAGCTCTTTCTTCAGACAGTTCCTCTATGATC 144
 Db 257 TTTCATTTTGACAGAGCTGAAGTATTTTCTTACACATTTCTCTTAAATTC 201

RESULT 6
 T66986/c
 ID T66986 standard; cDNA to mRNA; 1785 BP.
 AC T66986;
 DT 30-JUL-1997 (first entry)
 DE Human brain-specific adaptor FC99 cDNA.
 DE Brain-specific adaptor; FC99; neuron; signal transduction;
 KW neurological disease; Alzheimer's disease; Parkinson's disease;
 KW diagnosis; therapy; tyrosine kinase; ds.
 OS Homo sapiens.
 SN EP-773281-A2.
 PD 14-MAY-1997.
 PF 11-OCT-1996; 116333
 PR 13-OCT-1995; JP-263988.
 PR 12-DEC-1995; JP-323069.
 PR 29-FEB-1996; JP-069265.
 PR 24-JUL-1996; JP-212973.
 PA (SUME) SUMITOMO ELECTRIC IND CO.
 PI Nakamura T;
 DR WPI: 97-261318/24.
 DR P-5D8; W15256.
 PT Human and rat-derived brain-specific adaptor polypeptide FC99 - is
 PT involved in neuron signal transduction and is useful in diagnosis
 PT and treatment of neurological diseases e.g. Alzheimer's dementia
 PS Claim 4; Page 22-23; 42pp; English.
 CC A cDNA clone (T66986) codes for FC99 (W15256), a novel brain-
 CC specific adaptor molecule involved in neuron signalling pathways.
 CC It was obt'd. from a normalised cDNA library prep'd. from human
 CC cerebrum-derived mRNA. 5'-terminal sequences of plasmid inserts
 CC in E. coli transformants were sequenced and translated amino acid
 CC sequences were compared with a protein database for homology. Over
 CC 500 plasmids were analysed, and one plasmid, designated FC99 (FC -
 CC forebrain cortex), was selected. Sequencing revealed alternative,
 CC in-frame initiation points of translation, providing coding
 CC sequences of 1425 bases (T66983) and 1785 bases (T66986). FC99
 CC nucleic acids (including antisense) and polypeptides can be used to
 CC study phosphorylated tyrosine-mediated intracellular signalling
 CC pathways in diagnosis and treatment of neurological diseases and
 CC to test diagnosis/treatment methods or reagents.
 SQ Sequence 1785 BP; 412 A; 349 C; 522 G; 302 T;
 Query Match 16.0%; Score 28.8; DB 1; Length 1785;
 Best Local Similarity 62.5%; Pred. No. 2.9;
 Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 15 CCACATNTGGGCTGGGCTGCCCTTCTCTCTCTTGTGACCTCTTGGCAGCTCAGATG 74
 Db 419 CCCCGAGGGGGCTGGGCGGCGCTCTCGCGGGCGGCGCTCTCTCTCTCTCTCTCT 360
 QY 75 GAACAGGGCGCG 86
 Db 359 GCGGGGCGCGG 348
 RESULT 7
 Q71243
 ID Q71243 standard; cDNA; 1554 BP.
 AC Q71243;
 DT 22-MAR-1995 (first entry)
 DE Serine protease for fusion protein cleavage.
 DE Serine protease; Factor-Xa; recognition site;
 KW fusion protein cleavage; protein folding; ds.
 OS Bos taurus.
 SN Key
 FT cds
 FT key
 FT /tag= a
 FT mat_peptide
 FT /tag= b
 FT W09418227-A.
 PD 18-AUG-1994.

Db 64 GCAGCAGGCGCCTGGCAGACCCCTGGCCCCCTCTGGCCCATCTTCGG 15

RESULT 9

Q05106

Q05106 standard; DNA; 2259 BP.

ID Q05106;

AC 02-NOV-1990 (first entry)

AT Sequence encoding tumour-associated antigen GA733-1.

CT Pancreatic carcinoma; GA733-1; cancer; ds.

DE Homo sapiens.

ES

OS

QH Key

FT CDS Location/Qualifiers

FT 307..1275

FT /*tag= a

FT

PN EP-376746-A.

PD 4-JUL-1990.

PP 29-DEC-1989; 313687.

PR 29-DEC-1988; US-291583.

PA (WIST-) Wistar Inst.

PI Linnenbach A;

DR WPI: 90-203091/27.

DR P-FSDB; R03711.

PT Tumor-associated antigen, GA733-2 -

PT Expressed in pancreatic carcinoma cells, used for tumor

PT Immuno:therapy.

PS Claim 2; Fig 2A; 12pp; English.

CC The peptide is produced by tumor cells, and Abs raised to the

CC peptide may be used in diagnosis and therapy of human tumours.

SQ Sequence 2259 BP; 460 A; 648 C; 639 G; 512 T;

Query Match 15.2%; Score 27.4; DB 1; Length 2259;

Best Local Similarity 54.5%; Pred. No. 9.1;

Matches 55; Conservative 0; Mismatches 46; Indels 0; Gaps

QY 14 GCCACTATGGGCTGTGGCGTGGCCCTTGTCTCTCTTGACCTCTGGCAGCTCAGAT 73

Db 1144 GTCATCGTGGTGTGGGTGGCGCTCTGTCCGGCATGGCGTCTGTGGTATCCCAAC 1203

QY 74 GCACAGGCGCGGTATGACTTGGCACTGAAGTGAAGGA 114

Db 1204 CGGAGAAATCGGGGAAGTACAAAGAAGGTGGAGATCAGGA 1244

RESULT 10

T49318

ID T49318 standard; cDNA; 2122 BP.

AC T49318;

AT 30-MAR-1997 (first entry)

DE Wnt-10b gene.

KW Human; Wnt-10b; fetal bone marrow stroma cell; growth factor;

KW antibody; antisense; hematopoietic stem cell culture;

KW gene therapy; ex vivo; diagnostic; ss.

OS Homo sapiens.

QH Key

FT 5'utr Location/Qualifiers

FT 1..193

FT /*tag= a

FT 194..1363

FT /*tag= b

FT /product= Wnt-10b protein

FT 194..256

FT /*tag= c

FT 257..1360

FT /*tag= d

FT /product= Mature Wnt-10b (claim 19)

FT 1364..2122

FT /*tag= e

PN W09640910-A1.

PD 19-DEC-1996.

PP 06-JUN-1996; E02455.

PR 07-JUN-1995; US-485449.

PA (SANO) SANDOZ LTD.

PA (SANO) SANDOZ PAT. GMBH.

PA (SANO) SANDOZ-BEFLINDUNGEN VERW GMBH.
 PA (SYST-) SYSTEMIX INC.
 PI Van Den Berg DJ;
 DR WPI; 97-052317/05.
 DR P-PSDB; W08928.
 PT Expression vectors and methods for cloning N-terminal signal
 PT sequences - used for selection, expression and isolation of new
 PT signal sequences
 PS Claim 3; Page 28-30; 46pp; English.
 CC The sequence encodes human growth factor Wnt-10b, and has been
 CC isolated from a fetal bone marrow stroma cell cDNA library by PCR
 CC using consensus primers corresponding to conserved sequences within
 CC the Wnt gene family, followed by use of the amplified fragment as a
 CC probe and chromosome walking for isolation of the full-length
 CC sequence. A truncated sequence, Wnt-10b-delta, is given in 149319
 CC (claim 3). A plasmid containing the Wnt-10b cDNA has been deposited
 CC as ATCC 97208. Antibodies and antisense sequences corresponding to
 CC Wnt-10b may be used as pharmaceuticals, and Wnt-10b may be used in
 CC a culture medium for ex vivo haematopoietic stem cell culture for
 CC use e.g. in gene therapy. Fragments of the gene may be used as
 CC diagnostic probes or primers.
 SQ Sequence 2122 BP; 604 C; 635 G; 468 T;

Query Match 15.1%; Score 27.2; DB 1; Length 2122;
 Best Local Similarity 51.7%; Pred. No. 10;
 Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 15 CCACATGATGGTCTGGGCTGCCCTGTCCTCCCTCTGACCTCTTGGCAGCTCACATG 74
 Db 715 CCACTCTCTGCGACCGCTGGCTGGCTCAAGCCCGCCGCCAGGACACATG 774
 QY 75 GAACAGGCGCGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTCTGACAAAGTTCT 134
 Db 775 GGAATGGGTGGCTGTAACCATGACATGACTTTGGAGAGAGTCTCTCGGATTTCT 834

RESULT 11
 ID Q83695 standard; DNA; 2703 BP.
 AC Q83695;
 DT 10-OCT-1995 (first entry)
 DE New TGF-beta family member - MP-52 DNA sequence.
 KW Transforming growth factor-beta family; mitogenic; differentiation;
 KW treatment; prevention; disease; bone; cartilage; connective tissue;
 KW skin; mucosa; epithelium; dental tissue; wound healing; osteoporosis;
 KW tissue regeneration; arthritis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 640..2145
 FT /*tag= a
 FT /product= MP-52 TGF-beta propeptide
 FT mat_peptide 1783..2142
 FT /*tag= b
 FT /codon_start= 1783
 FN W09504819-A.
 PD 16-FEB-1995.
 PF 09-AUG-1994; E02630.
 PR 10-AUG-1993; DE-326829.
 PR 25-MAY-1994; DE-418222.
 PR 09-JUN-1994; DE-420157.
 PA (BIOP-) BIOPHARM GFS BIOTECHNOLOGISCHEN ENTWICKL.
 PI Hotten G, Neidhardt H, Paulista M, Hoetten G;
 DR WPI; 95-090897/12.
 DR P-PSDB; R69600.
 CC New DNA encoding a new member of the TGF beta family - and
 CC related vectors, host cells etc., has mitogenic and
 CC differentiation inducing activity, e.g. for treating or
 CC preventing diseases of bone and cartilage etc.
 PS Claim 1; Page 34-35; 51pp; German.
 CC The nucleotide sequence of a novel member of the transforming growth
 CC factor-beta (TGF-beta) family named MP-52. The gene encodes a protein
 CC of 501 amino acids (AA). The protein, or at least the mature protein,

CC has mitogenic and/or differentiation inducing properties useful in
 CC the treatment or prevention of diseases of bone, cartilage, connective
 CC tissue, skin, mucosa, epithelium or dental tissue. The protein can also
 CC be used for wound healing and tissue regeneration e.g. in osteoporosis
 CC and arthritis.
 SQ Sequence 2703 BP; 602 A; 758 C; 784 G; 559 T;

Query Match 14.9%; Score 26.8; DB 1; Length 2703;
 Best Local Similarity 64.5%; Pred. No. 15;
 Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 13 GCCCATGATGGGTGGGCTGCCCTGTCCTCCCTCTGACCTCTTGGCAGCTCAC 72
 Db 2449 GCCCACCATTCTCTCACCCTGGGCTCTTCAGCTCTGGACTCTCTTAGCACCTCTCA 2508
 QY 73 TG 74
 Db 2509 GG 2510

RESULT 12
 ID T31601 standard; DNA; 2341 BP.
 AC T31601;
 DT 23-OCT-1996 (first entry)
 DE Cartilage-derived morphogenetic protein-1 gene.
 KW Human; cartilage-derived morphogenetic protein-1; CDMP-1;
 KW articular cartilage; chondrogenic; vulnery; implantation;
 KW chondromalacia; osteoarthritis; therapy; joint repair; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_feature 1..381
 FT /*tag= a
 FT /*note= "Genomic DNA sequence"
 FT cds 265..1770
 FT /*tag= b
 FT /product= Cartilage-derived morphogenetic protein-1
 FT signal_peptide 268..321
 FT /*tag= c
 FT mat_peptide 322..1767
 FT /*tag= d
 FT /product= Pro-CDMP-1 protein
 FT misc_feature 382..2341
 FT /*tag= e
 FT /*note= "cDNA sequence"
 FT mat_peptide 1408..1767
 FT /*tag= f
 FT /product= Mature C-terminal domain
 FT misc_feature 1501..1620
 FT /*tag= g
 FT /*note= "PCR-amplified probe used to screen cDNA
 FT library"
 FN W09614335-A1.
 PD 17-NAY-1996.
 PF 07-NOV-1994; WO-U12814.
 PR 07-NOV-1994; WO-U12814.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chang SC, Luyten FP, Moos W;
 DR WPI; 96-251714/25.
 DR P-PSDB; R95635.
 CC New purified cartilage extracts and proteins - used to stimulate the
 CC development and repair of cartilage in vivo.
 PS Claim 10; Fig 1; 34pp; English.
 CC The sequence encodes human articular cartilage-derived morphogenetic
 CC protein-1 (CDMP-1). Primers T13603-06, constructed from conserved
 CC motifs from bone morphogenetic protein, have been used in reverse
 CC transcription-polymerase chain reaction amplification of a cattle
 CC articular chondrocyte cDNA library, generating products of 120 and
 CC 280 bp, which are then used to screen a human articular cartilage
 CC cDNA library. A positive clone lacking the N-terminal region has
 CC been isolated, and used to obtain a full-length genomic clone.
 CC CDMP-1 is present in a purified cartilage extract (claimed) which

Qy 8 GCACTGGCCACTATGGGGTCTGGGCTGCCCCCTCTCCTCTCTTGACCTCCTTGGCAGC 67

Db 128 GCGGTGGCCAGGCCGCGTCACTGCTGGCGGCCCAACATGGCCCTGTGCCCGGTCTGTGGC 69

QY 58 TCACATGGACAGGCGCGGTATGACTTTGCACTG 103
 Db 58 AGAAATGGTCTTCGCCGCGCGTGCAGATTGCTACAG 33

RESULT 2

PCT-US95-04636-7/c
 ; Sequence 7, Application PC/TUS9504636
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
 ; NUMBER OF SEQUENCES: 10
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII(text)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/04636
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/346,147
 ; FILING DATE: 29-NOV-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/306,511
 ; FILING DATE: 14-SEP-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/248,812
 ; FILING DATE: 25-MAY-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/227,371
 ; FILING DATE: 14-APR-1994
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 580 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: both
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 91...480
 PCT-US95-04636-7

Query Match 16.9%; Score 30.4; DB 6; Length 580;
 Best Local Similarity 57.3%; Pred. No. 0.12;
 Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
 QY 8 GCAGTGGCCACTATGGGCTGCGGCTGCTCTTCTCTTGTGACCTCTCTTGGGAGC 67
 Db 128 GCGGTGGCGAGCGCGGTGCTGCTGCGGCGGCAACATGCGCTTGTCCCGGTGTGTC 69
 QY 68 TCACATGGACAGGCGCGGTATGACTTTGCACTG 103
 Db 68 AGAAATGGTCTTCGCCGCGCGTGCAGATTGCTACAG 33

RESULT 3

US-08-469-486-1
 ; Sequence 1, Application US/08469486
 ; Patent No. 5739281
 ; GENERAL INFORMATION:
 ; APPLICANT: Thøgersen, Hans Christian
 ; APPLICANT: Hollet, Thor Las
 ; APPLICANT: Etzerodt, Michael
 ; TITLE OF INVENTION: Improved method for the refolding of
 ; TITLE OF INVENTION: Proteins
 ; NUMBER OF SEQUENCES: 58
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version
 ; SOFTWARE: #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/469,486
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/192,060
 ; FILING DATE: February 4, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Paul T. Clark
 ; REGISTRATION NUMBER: 30,162
 ; REFERENCE/DOCKET NUMBER: 06363/002001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617 542 5070
 ; TELEFAX: 617 542 8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1554 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: YES
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Bos taurus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 76..1551
 US-08-469-486-1

Query Match 15.9%; Score 28.6; DB 2; Length 1554;
 Best Local Similarity 49.0%; Pred. No. 0.77;
 Matches 76; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
 QY 15 CCACATATGGGGTCTGGGCTGCGGCTCTCTCTCTTGTGACCTCTCTTGGCAGCTCACATG 74
 Db 71 CCACATGCGGCGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 130
 QY 75 GAACAGGCGCGGTATGACTTTGCACTGAAGCTGAAGGAGTCTTTTCTGACAAAGTCTCT 134
 Db 131 GGCGGCGGGGAGCGGTCT 190
 QY 135 CCAATGAGTCCAGCTTCTCTGGAATTCCTGAAAG 169
 Db 191 GCAGGCGCAACTCATCTCTTGAGGAGGTGAACAG 225

RESULT 4

US-08-469-658-1
 ; Sequence 1, Application US/08469658
 ; Patent No. 5917018
 ; GENERAL INFORMATION:
 ; APPLICANT: Thøgersen, Hans Christian
 ; APPLICANT: Hollet, Thor Las
 ; APPLICANT: Etzerodt, Michael
 ; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOOLDING OF
 ; TITLE OF INVENTION: PROTEINS
 ; NUMBER OF SEQUENCES: 58
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.


```
RESULT 7
US-08-485-449-1
; Sequence 1, Application US/08485449
; Patent No. 5824789
; GENERAL INFORMATION:
; APPLICANT: VANDENBERG, DAVID
; TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,449
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: KOSKI, ANTOINETTE F.
; REGISTRATION NUMBER: 34,202
; REFERENCE/DOCKET NUMBER: 20296-20035.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 2122 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 194..1360
US-08-485-449-1

Query Match 15.1%; Score 27.2; DB 2; Length 2122;
Best Local Similarity 51.7%; Pred. No. 2.7;
Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 15 CCACATATGGGCTGGGCTGGCCCTTCTCTCTCTTGTGACCCCTCTTGGCAGCTCACATG 74
Db 715 CCACCTCTCTGCCAGCCCTGGCCCTGGCTCAAGCCCGCCCTGGCCCGCCAGCACATG 774

QY 75 GAACAGGCGGGTATGACTTTGCACTGAAGCTGAAGCTGTTTCTGACAAAGTTCTT 134
Db 775 GGAATGGGCTGCTTAACCATGACATGGACTTTGGAGAGAGTTCTCTCGGGAATTTCTT 834

RESULT 8
US-08-288-508C-1
; Sequence 1, Application US/08288508C
; Patent No. 5994094
; GENERAL INFORMATION:
; APPLICANT: H tten, Gertrud
; APPLICANT: Neidhardt, Helge
; APPLICANT: Paulista, Michael
; TITLE OF INVENTION: NEW GROWTH/DIFFERENTIATING FACTOR OF
; TITLE OF INVENTION: THE TGF- FAMILY
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikolaio, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
```

```
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,508C
; FILING DATE: 10-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 43 26 829.3
; FILING DATE: 10-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 18 222.8
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 20 157.5
; FILING DATE: 09-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: JAHNS, Kristina M.
; REGISTRATION NUMBER: P-41,092
; REFERENCE/DOCKET NUMBER: P584-4019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 2703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-288-508C-1

Query Match 14.9%; Score 26.8; DB 4; Length 2703;
Best Local Similarity 64.5%; Pred. No. 4.2;
Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 13 GGCCATATGGGGCTGGGCTGGCCCTTCTCTCTCTTGTGACCCCTCTTGGCAGCTCAC 72
Db 2449 GCCCACCATTCTCTCTCACCTGGGCGCTTCTCAGCCCTCTGGACTCTCTAAGCAGCTCTCA 2508

OY 73 TG 74
Db 2509 GG 2510

RESULT 9
US-08-937-466-1/c
; Sequence 1, Application US/08937466
; Patent No. 5846779
; GENERAL INFORMATION:
; APPLICANT: Zhang, Ning
; APPLICANT: Amara, M. Catherine
; APPLICANT: Chen, Jin-Long
; TITLE OF INVENTION: UCP3 Genes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/937,466
: FILING DATE:
:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: T97-009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 343-4341
: TELEFAX: (650) 343-4342
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2782 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: US-08-937-466-1
:
: Query Match 14.8%; Score 26.6; DB 3; Length 2782;
: Best Local Similarity 53.3%; Pred. No. 5;
: Matches 56; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
:
: QY 75 GACAGGGCCGGTATGACTTTGGCACTGAGCTGAGGAGTCTTTCTTGACAGTTCCT 134
:
: Db 2101 GAAACCTCCGAGAGAGCGGAGGAGGAGCTGTCGTCTTCCTCCCTAAGGTGGAG 2042
:
: QY 135 CCTATGAGTCCAGCTTCCTCGAATTCCTTGAAGAGCTCTGCGCTCC 179
:
: Db 2041 ATTCGAGTCTGCTGCTTTGAACGTGATGGACAGGCGTGGCCCC 1997
:
: RESULT 10
: US-09-172-528-1/c
: Sequence 1, Application US/09172528
: Patent No. 5952469
: GENERAL INFORMATION:
: APPLICANT: Zhang, Ning
: APPLICANT: Amaral, M. Catherine
: APPLICANT: Chen, Jin-Long
: TITLE OF INVENTION: UCP3 Genes
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 75 DENISE DRIVE
: CITY: HILLSBOROUGH
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94010
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/172,528
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/937,466
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: T97-009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 343-4341
: TELEFAX: (650) 343-4342
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 22, 2000, 05:09:44 ; Search time 802.03 Seconds
(without alignments)
847.374 Million cell updates/sec

Title: US-09-092-296-3
Perfect score: 180
Sequence: 1 CAGGAGCGCAGTGGCCACTA.....CTTGAAAGCTCTGCGCTCT 180

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*

- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_gss1:*
- 80: gb_gss2:*
- 81: gb_gss3:*
- 82: gb_gss4:*
- 83: em_gss1:*
- 84: em_gss2:*
- 85: em_gss3:*
- 86: em_gss4:*
- 87: gb_gss5:*
- 88: gb_gss6:*
- 89: gb_gss7:*
- 90: gb_gss8:*
- 91: gb_gss9:*
- 92: em_gss5:*
- 93: em_gss6:*
- 94: em_gss7:*
- 95: em_gss8:*
- 96: em_gss9:*
- 97: em_gss10:*
- 98: em_gss11:*
- 99: gb_gss10:*
- 100: gb_gss11:*
- 101: em_gss12:*
- 102: gb_gss12:*
- 103: gb_gss13:*
- 104: gb_gss14:*
- 105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	100.4	55.8	404	61	A1857998
C 2	74.2	41.2	552	82	AQ18761
C 3	55.6	30.3	328	42	A1136523
C 4	33.4	18.6	568	100	AQ290553
C 5	32.4	18.0	537	64	AW078074
C 6	32.4	18.0	523	69	AW128453
C 7	31.4	17.4	339	21	F08745

8 31 17.2 545 23 T41524
 9 30.8 17.1 427 30 AA219322
 10 30.8 17.1 589 51 AT730494
 11 30.4 16.9 392 26 W70374
 12 30.4 16.9 742 27 Z78408
 13 30.4 16.9 290 33 AA387480
 14 30.4 16.9 524 43 AT179983
 15 30.4 16.9 445 87 AQ768084
 16 30 16.7 585 63 AT1998636
 17 29.8 16.6 505 20 M89011
 18 29.8 16.6 401 89 AQ904556
 19 29.6 16.4 547 23 T41523
 20 29.6 16.4 573 23 T41535
 21 29.4 16.3 293 37 AA668276
 22 29.4 16.3 344 37 AA716892
 23 29.4 16.3 323 38 AA717760
 24 29.4 16.3 286 38 AA762265
 25 29.4 16.3 421 39 AA839043
 26 29.4 16.3 400 46 AT463192
 27 29.4 16.3 517 61 AT839539
 28 29.4 16.3 530 62 AT881867
 29 29.4 16.3 408 64 AL117984
 30 29.4 16.3 603 64 AW66022
 31 29.4 16.3 455 64 AW66263
 32 29.4 16.3 446 104 AO527267
 33 29.4 16.3 382 105 AQ613185
 34 29.2 16.2 601 51 AT727080
 35 29.2 16.2 592 51 AT730002
 36 29.2 16.2 677 100 AQ316270
 37 29.2 16.2 452 105 AQ647097
 38 29.2 16.2 499 105 AQ647978
 39 29 16.1 420 38 AA774353
 40 29 16.1 685 103 AQ473733
 41 28.8 16.0 266 21 R09726
 42 28.8 16.0 574 51 AT730173
 43 28.8 16.0 847 99 AQ258186
 44 28.8 16.0 457 103 AQ500124
 45 28.6 15.9 536 23 T41525

ALIGNMENTS

RESULT 1
 LOCUS AI857998/c
 DEFINITION w169b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408041 3',
 mRNA sequence. EST 26-AUG-1999
 ACCESSION AI857998.1 GI:5511614
 VERSION AI857998
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 404)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/obrp/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 395.
 FEATURES
 Location/Qualifiers
 1..404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI_CGAP_Lu19"
 /tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 97 a 105 c 117 g 84 t 1 others
 ORIGIN
 Query Match 55.8%; Score 100.4; DB 61; Length 404;
 Best Local Similarity 98.1%; Pred. No. 6.7e-21;
 Matches 101; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 78 CAGGCGCGGTGACTTTCACACTGAGCTGAGCTGAGCTGCTTTCTGACAAAGTCTCTCT 137
 Db 311 CAGGCGCGGTGACTTTCACACTGAGCTGAGCTGAGCTGCTTTCTGACAAATCTCTCT 252
 QY 138 ATGAGTCCAGCTTCCTGGAATTCGTTGAAAGCTCTGCTCTCT 180
 Db 251 ATGAGTCCAGCTTCCTGGAATTCGTTGAAAGCTCTGCTCTCT 209
 RESULT 2
 AQ718761/c
 LOCUS AQ718761 552 bp DNA GSS 13-JUL-1999
 DEFINITION HS_5511_B2_F09_T7A_RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate-1087 Col-18 Row=L, genomic survey sequence.
 ACCESSION AQ718761
 VERSION AQ718761.1 GI:5468077
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1087 row: L column: 18

```

Seq primer: T7
Class: BAC ends
High quality sequence stop: 552.
FEATURES
    source
        1..552
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate-1087 Col-18 Row-L"
            /clone_lib="RPC1-11 Human Male BAC Library"
            /sex="male"
            /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      141 a   153 c   124 g   117 t   17 others
ORIGIN
Query Match      41.2%; Score 74.2; DB 82; Length 552;
Best Local Similarity 85.4%; Pred. No. 8.7e-13;
Matches 82; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 76 AACAGGCGGGTATGACTTTGCACTGAAGCTGAAGGAGTCTTTCTGACAGTTCTCTC 135
Db 474 ACGATGGGAGGTATGACTTTGCANCTGAAGCTGAGAGGTCTATTAGACAAATCTCTC 415

QY 136 CTATGATCCAGCTTCCTGGAAATGCTTGAAGAGCT 171
Db 414 TTATGAGTCAGCTTCTCTGGAATGCTTGAAGAGGT 379

RESULT 3
LOCUS      A1136523/c      328 bp      mRNA      EST      11-FEB-1999
DEFINITION UI-R-C2p-ng-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
ACCESSION  A1136523
VERSION     A1136523.1 GI:3637300
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE     1 (bases 1 to 328)
JOURNAL   Bonaldo,M.F., Lennon,G. and Soares,M.B.
MEDLINE   Normalization and subtraction: two approaches to facilitate gene
COMMENT   discovery
          Genome Res. 6 (9), 791-806 (1996)
          97044477
          On Jan 19, 1998 this sequence version replaced gi:2150222.
          Contact: Soares, MB
          Program for Rat Gene Discovery and Mapping
          University of Iowa
          451 Eckstein Medical Research Building Iowa City, IA 52242, USA
          Tel: 319 335 8250
          Fax: 319 335 9565
          Email: msoares@blue.weeg.uiowa.edu
          The sequence tag present in the cDNA between the NotI site and the
          oligo-dr track served to identify it as a clone from the normalized
          adult lung library. cDNA Library Preparation: M. Fatima Bonaldo,
          Ph.D. Clone distribution: clones will be available through Research
          Genetics
          Seq primer: M13 Forward.
          Location/Qualifiers
            1..328
              /organism="Rattus norvegicus"
              /strain="Sprague-Dawley"
              /db_xref="taxon:10116"
              /clone="UI-R-C2p-ng-e-02-0-UI"
              /clone_lib="UI-R-C2p"
              /dev_stage="adult"
              /lab_host="DH10B (Life Technologies)"
              /note="Vector: pT73D-Pac (Pharmacia) with a modified
              polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p
              library is a subtracted library derived from the UI-R-C1
              library, which is a subtracted library derived from the

```

```

UI-R-C0 library. The UI-R-C0 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2p) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2p library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996)."
BASE COUNT      62 a   77 c   98 g   91 t
ORIGIN
Query Match      30.9%; Score 55.6; DB 42; Length 328;
Best Local Similarity 66.9%; Pred. No. 3.8e-07;
Matches 79; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 63 GCAGCTCACATGACAGGCGGGTATGACTTTGCACTGAAGCTGAAGGAGTCTTTTC 122
Db 327 GCAGCTCATCTGAGGGCCACAAAGTGACTTTGCACTGAAGGAGTCTTTTC 268

QY 123 TGACAGATTCTTCCTATGATGCCAGCTTCTCTGGAATGCTTGAAGGAGTCTTCCTCT 180
Db 267 AGCCAGAGACCTCCCAAGACTCGGGCTTTCTGGACATGCTCCAAAGAGATCTGCCTCT 210

RESULT 4
LOCUS      AQ290553      568 bp      DNA      GSS      03-DEC-1998
DEFINITION nbx50037C20f CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION  AQ290553
VERSION     AQ290553.1 GI:3951915
KEYWORDS   GSS.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa.
REFERENCE  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
TITLE     Poaceae; Oryza.
JOURNAL   1 (bases 1 to 568)
COMMENT   Wing,R.A. and Dean,R.A.
          A BAC End Sequencing Framework to Sequence the Rice Genome
          Unpublished (1998)
          Contact: Wing RA
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 7288
          Fax: 864 656 4293
          Email: rwing@clemson.edu
          Seq primer: TAATAGCACTCACTATAGGG
          Class: BAC ends
          High quality sequence start: 35
          High quality sequence stop: 317.
          Location/Qualifiers
            1..568
              /organism="Oryza sativa"
              /strain="Japonica"
              /cultivar="Nipponbare"
              /db_xref="taxon:4530"
              /clone="nbx50037C20f"
              /clone_lib="CUGI Rice BAC Library"
              /tissue_type="Leaf"

```



```

Db 243 TGTCCTGACCCCATCTACTCTCTCTGGGACTTCTAGTCACACAGGCTGGGCTGG 302
QY 93 CTTTGCACTGAAG 106
Db 303 ACTTCANCTGNAG 315

RESULT 13
AA387480 290 bp mRNA EST 23-APR-1997
LOCUS v83d07.r1 Ko mouse embryo 11 5dpc Mus musculus cDNA clone
DEFINITION IMAGE:789613 5', mRNA sequence.
ACCESSION AA387480
VERSION AA387480.1 GI:2040672
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 290)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1317023.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:474021
High quality sequence stop: 81.
FEATURES
Source
1..290
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:789613"
/clone_lib="Ko mouse embryo 11 5dpc"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="11.5dpc"
/lab_host="DB10B"
/Note="Organ: Embryo; Vector: pSPORT1; Site_1: SalI;
Site_2: NotI; Total RNAs were extracted from 11.5 dpc
embryos (excluding placenta and yolk sac). The
double-stranded cDNA was synthesized with an oligo (dT)-1
primer GAGAGAGACTAGTCTAGATCGAGCGCGGCTTTTTTTTTTTT
3'. The cDNAs were ligated to LL-Sal3a: 5'
GCTATTGAGTGGACTATCC 3' and LL-Sal3B: 5'
GGATGCGAGCTCAAT 3'. The cDNAs were size-selected and
amplified by long-range PCR using Ex taq polymerase for 18
cycles. The PCR-amplifiable cDNA mixture went through
one round of equalization and was digested with SalI/NotI
and cloned into the SalI/NotI sites of the pSPORT1
plasmid vector (Life Technologies). The library was
constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong
Wang."
BASE COUNT 71 a 81 c 82 g 56 t
ORIGIN

Query Match 16.9%; Score 30.4; DB 33; Length 290;
Best Local Similarity 61.2%; Pred. NO. 20;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

AA387480 290 bp mRNA EST 23-APR-1997
LOCUS v83d07.r1 Ko mouse embryo 11 5dpc Mus musculus cDNA clone
DEFINITION IMAGE:789613 5', mRNA sequence.
ACCESSION AA387480
VERSION AA387480.1 GI:2040672
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 290)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1317023.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:474021
High quality sequence stop: 81.
FEATURES
Source
1..290
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:789613"
/clone_lib="Ko mouse embryo 11 5dpc"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="11.5dpc"
/lab_host="DB10B"
/Note="Organ: Embryo; Vector: pSPORT1; Site_1: SalI;
Site_2: NotI; Total RNAs were extracted from 11.5 dpc
embryos (excluding placenta and yolk sac). The
double-stranded cDNA was synthesized with an oligo (dT)-1
primer GAGAGAGACTAGTCTAGATCGAGCGCGGCTTTTTTTTTTTT
3'. The cDNAs were ligated to LL-Sal3a: 5'
GCTATTGAGTGGACTATCC 3' and LL-Sal3B: 5'
GGATGCGAGCTCAAT 3'. The cDNAs were size-selected and
amplified by long-range PCR using Ex taq polymerase for 18
cycles. The PCR-amplifiable cDNA mixture went through
one round of equalization and was digested with SalI/NotI
and cloned into the SalI/NotI sites of the pSPORT1
plasmid vector (Life Technologies). The library was
constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong
Wang."
BASE COUNT 71 a 81 c 82 g 56 t
ORIGIN

```

```

QY 2 AGAGCGCAGTGGCCACTATGCGGCTCGCCCTTGCTCTCTTGACCCCTCTT 61
Db 123 AGATCGCGGGCGCGCTCATAGGACGCTGCTGCGCCCTTGCTCATCGGGGCCATCCTC 182
QY 62 GGCAGCTCACATGGAACAGG 81
Db 183 TTCTGCTGTACAGGAACG 202

RESULT 14
AA179983 524 bp mRNA EST 20-JAN-1999
LOCUS EST223714 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone
DEFINITION RSPC049 3' end, mRNA sequence.
ACCESSION AA179983
VERSION AA179983.1 GI:3730621
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 524)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..524
/organism="Rattus sp."
/db_xref="ATCC (inhost):2035352"
/db_xref="taxon:10118"
/clone="RSPC049"
/clone_lib="Normalized rat spleen, Bento Soares"
/Note="Organ: spleen; Vector: p7T3Fac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 129 a 124 c 141 g 130 t
ORIGIN

Query Match 16.9%; Score 30.4; DB 43; Length 524;
Best Local Similarity 61.2%; Pred. NO. 26;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 57 TCCTTGGCAGCTCACATGGAACAGGCGCGGCTATGCTTTGCAACTGAAGCTGAAGGAGT 116
Db 443 TCCTTGGCACCACGATCCAACTCTCGGGTTCGCTGTGAACGAACCTGGGAGCT 502
QY 117 CTTTCTTCACAGTTCTCC 136
Db 503 CTACCTGACTAGTTCACCC 522

RESULT 15
AAQ768084 445 bp DNA GSS 28-JUL-1999
LOCUS HS_3037.B1.E11.MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-3037 Col-21 Row-J, genomic survey
sequence.
ACCESSION AAQ768084
VERSION AAQ768084.1 GI:5646200
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

```

Wed Mar 22 12:07:49 2000

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 445)
Mahairas.G.G., Wallace.J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 3097 row: J column: 21

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 445.

FEATURES

source

1..445

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="plate-3097 Col-21 Row-J"

/clone_lib="CII Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B,

BASE COUNT 139 a 114 c 70 g 121 t 1 others

ORIGIN

Query Match 16.9%; Score 30.4; DB 87; Length 445;
Best Local Similarity 57.2%; Pred. No. 24;
Matches 43; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 35 CCCCTTGTCTCTTGTACCTCCTTGGCAGCTCAGTGAACAGGCGGGGTATGACT 94

Db 101 CCCCTTGTCTCTGTATGCCCCCTTATTACTATTCTCATGAGACAGCTCCCGCTATGACT 160

QY 95 TTGC 98

Db 161 ATGC 164

Search completed: March 22, 2000, 05:09:46
Job time: 1367 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2000, 14:39:02 ; Search time 479.47 Seconds
(without alignments)
-1450.209 Million cell updates/sec

Title: US-09-092-296-2
Perfect score: 229
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCCTTCAGGGACCA 229

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl.*

Word size : 0

Number of hits that pass the threshold : 1642386

1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_om.*
5: gb_om.*
6: gb_ph.*
7: gb_ph.*
8: gb_ph.*
9: gb_ph.*
10: gb_ph.*
11: gb_ph.*
12: gb_ph.*
13: gb_ph.*
14: gb_ph.*
15: gb_ph.*
16: gb_ph.*
17: gb_ph.*
18: gb_ph.*
19: gb_ph.*
20: gb_ph.*
21: gb_ph.*
22: gb_ph.*
23: gb_ph.*
24: gb_ph.*
25: gb_ph.*
26: gb_ph.*
27: gb_ph.*
28: gb_ph.*
29: gb_ph.*
30: gb_ph.*
31: gb_ph.*
32: gb_ph.*
33: gb_ph.*
34: gb_ph.*
35: gb_ph.*
36: gb_ph.*
37: gb_ph.*
38: gb_ph.*
39: gb_ph.*
40: gb_ph.*
41: gb_ph.*
42: gb_ph.*
43: gb_ph.*
44: gb_ph.*
45: gb_ph.*
46: gb_ph.*
47: gb_ph.*
48: gb_ph.*
49: gb_ph.*

50: gb_pl3.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	109.4	47.8	192650	10	AB023048	AB023048 Homo sapi
2	109.4	47.8	200000	10	AF000511	AF000511 Homo sapi
3	109.4	47.8	47323	11	AC005937	AC005937 Homo sapi
C 4	34.4	15.0	77322	10	HSDJ14C9	AL096774 Human DNA
C 5	34.4	15.0	211190	44	AC016716	AC016716 Homo sapi
C 6	34.4	15.0	191010	45	AC016764	AC016764 Homo sapi
7	34	14.8	1801	7	OSU25430	U25430 Oryza sativ
8	34	14.8	36676	11	HSJ858B16	AL096768 Human DNA
9	34	14.8	291288	42	AC008878	AC008878 Homo sapi
10	33.8	14.8	62181	44	AC016298	AC016298 Homo sapi
C 11	33.6	14.7	138379	11	HSJ773A18	AL049557 Human DNA
C 12	33.6	14.7	200000	41	AC006445	AC006445 Homo sapi
C 13	33.6	14.7	36065	43	AC014817	AC014817 Drosophil
C 14	33.4	14.6	106791	45	AC017080	AC017080 Homo sapi
C 15	32.6	14.2	5280	9	ABG11114	AB011114 Homo sapi
16	32.4	14.1	97580	11	HS591N18	AL031594 Human DNA
C 17	31.8	13.9	62091	32	DMR28018	AL121814 Drosophil
C 18	31.6	13.8	2803	2	TNU93354	U93354 Thermotoga
C 19	31.6	13.8	566	7	AAAL7A6	X78222 A.alternata
C 20	31.6	13.8	570	8	AAU87806	U87806 Alternaria
C 21	31.6	13.8	158986	40	AC008013	AC008013 Homo sapi
C 22	31.6	13.8	179622	43	AC009533	AC009533 Homo sapi
C 23	31.4	13.7	2745	3	EEU33171	U33171 Erinaceus e
C 24	31.4	13.7	39261	10	AC000387	AC000387 Genomic s
C 25	31.4	13.7	118067	10	HS163M9	AL021970 Homo sapi
C 26	31.4	13.7	303806	33	AL133353	AL133353 Homo sapi
C 27	31.4	13.7	154791	45	AC015478	AC015478 Homo sapi
C 28	31.2	13.6	506	3	CHIRONA	X71732 C.hircus in
C 29	31.2	13.6	177241	11	HS402G11	AL022328 Human DNA
C 30	31.2	13.6	78054	35	AC004296	AC004296 Drosophil
C 31	31.2	13.6	179532	42	AC010856	AC010856 Homo sapi
C 32	31.2	13.6	245767	42	AC011607	AC011607 Homo sapi
C 33	31.2	13.6	207043	44	AC011612	AC011612 Homo sapi
C 34	31	13.5	103150	8	ATF11111	AL079347 Arabidops
C 35	31	13.5	125502	8	ATT4L20	AL023094 Arabidops
C 36	31	13.5	42244	11	AC004660	AC004660 Homo sapi
C 37	31	13.5	39303	11	AC005595	AC005595 Homo sapi
C 38	31	13.5	190837	11	HS90L6	Z97353 Human DNA s
C 39	31	13.5	727	34	MITW17	L31623 Mitocoma c
C 40	31	13.5	186851	40	AC007297	AC007297 Homo sapi
C 41	31	13.5	97339	44	AC011039	AC011039 Homo sapi
C 42	31	13.5	208079	44	AC013391	AC013391 Homo sapi
C 43	30.8	13.4	14524	10	D86997	D86997 Homo sapien
C 44	30.8	13.4	46604	10	D88269	D88269 Homo sapien
C 45	30.8	13.4	100635	11	AC003104	AC003104 Homo sapi

ALIGNMENTS

RESULT 1
AB023048
LOCUS
DEFINITION
Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
clone:53L9, complete sequence.
ACCESSION
AB023048
VERSION
AB023048.1
KEYWORDS
Htg.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

1. (sites)
Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,I.,
Watanabe,K., Fukusumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Babram,S. and Inoko,H.
Molecular dynamics of MHC genes unraveled by sequence analysis of
the 1.796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
20027539
2 (bases 1 to 192650)
Shina,T. and Takishima,N.
Direct Submission
Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shina, Tokai University School of Medicine, Department of
Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail:tshina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)
Location/Qualifiers
1. .192650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="978SK"
/chromosome="6"
/clone="5319"
/map="6p21.3"
49862 a 44743 c 45833 g 52212 t

BASE COUNT
ORIGIN

Query Match 47.8%; Score 109.4; DB 10; Length 192650;
Best Local Similarity 95.0%; Pred. No. 2.9e-24;
Matches 113; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Oy 1 ACCGGGACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCATGATGGGTCTGGGCT 60
|||||
Db 177635 ACCGGGACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCATGATGGGTCTGGGCT 177694
Oy 61 GCCCCTTGCTCTCTGACCCCTCTTGGCAGCTCATATGACAGGCGCGGTATGA 119
|||||
Db 177695 GCCCCTTGCTCTCTGACCCCTCTTGGCAGCTCATATGACAGGCGCGGTATGA 177753
RESULT 2
AP000511 200000 bp DNA PRI 28-SEP-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
DEFINITION section 10/20.
ACCESSION AP000511
VERSION AP000511.1 GI:5926698
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM
Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Shina,S., Tamiya,G., Oka,A. and Inoko,H.
Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
Published Only in Database (1999) In press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,T.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www.alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS

Project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (<http://www.alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www.alis.tokyo.jst.go.jp.

FEATURES
source
1. .200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
108774. .108864
/standard_name="D6S2088"
/note="SHGC-12985:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735268"
115648. .115906
/standard_name="D6S1898"
/note="WI-9418:The location is between each flanking site
of PCR primers."
/db_xref="GDB:678272"
complement(115665. .115896)
/note="Cda0v10:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443382"
142030. .142139
/standard_name="D6S2203"
/note="SHGC-16870:The location is between each flanking
site of PCR primers."
/db_xref="GDB:741215"
complement(142085. .142322)
/note="RH18132:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4573021"
complement(142479. .142740)
/standard_name="D6S1851"
/note="SHGC-10808:The location is between each flanking
site of PCR primers."
/db_xref="GDB:675281"
complement(164525. .164665)
/note="SHGC-3064:The location is between each flanking
site of PCR primers."
/db_xref="GDB:1234116"
complement(169209. .169574)
/standard_name="D6S952"
/note="UT5233:The location is between each flanking site
of PCR primers."
/db_xref="GDB:313481"
BASE COUNT 52605 a 47531 c 49366 g 50498 t
ORIGIN

Query Match 47.8%; Score 109.4; DB 10; Length 200000;
Best Local Similarity 95.0%; Pred. No. 2.9e-24;
Matches 113; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Oy 1 ACCGGGACTTCAGTCTCTCTCCATCCAGGAGCGCAGTGGCCATGATGGGTCTGGGCT 60
|||||
Db 109962 ACCGGGACTTCAGTCTCTCTCCATCCAGGAGCGCAGTGGCCATGATGGGTCTGGGCT 110021
Oy 61 GCCCCTTGCTCTCTGACCCCTCTTGGCAGCTCATATGACAGGCGCGGTATGA 119
|||||
Db 110022 GCCCCTTGCTCTCTGACCCCTCTTGGCAGCTCATATGACAGGCGCGGTATGA 110080
RESULT 3
AC005937 47323 bp DNA PRI 05-NOV-1998
LOCUS Homo sapiens clone UWCC:370M23.002 from 6p21, complete sequence.
DEFINITION
ACCESSION AC005937
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.


```

Matches 113; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ACCGGACATTCAGTCTCTCCATCCAGGAGGCGAGTGCCACTATGGGCTCTGGGCT 60
|||||
Db 35121 ACCGGACATTCAGTCTCTCCATCCAGGAGGCGAGTGCCACTATGGGCTCTGGGCT 35180
|||||
QY 61 GCCCTTGTCTCTCTGACCTCTTGGCAGCTCACATGGAACAGCCGGGTAAGA 119
|||||
Db 35181 GCCCTTGTCTCTCTGACCTCTTGGCAGCTCACATGGAACAGGAGGCTAGA 35239
|||||

RESULT 4
HSDJ144C9/c 77322 bp DNA PRI 22-NOV-1999
LOCUS Human DNA sequence from clone RPI-144C9 on chromosome 1p34.3-36.11,
DEFINITION complete sequence.
ACCESSION AL096774
VERSION AL096774.9 GI:6465842
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 77322)
AUTHORS Hall, R.
TITLE Direct Submission
JOURNAL
COMMENT
On Nov 22, 1999 this sequence version replaced gi:6138775.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RPI-144C9 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCVPAC2
This sequence is the entire insert of clone RPI-144C9.
FEATURES
Source 1, 77322
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="p34.3-36.11"
/clone="RPI-144C9"
/clone_lib="RPCI-1"
complement(1710..2227)
/feature="match; GSS: Em:AQ355665"
complement(1780..2227)
/feature="match; GSS: Em:AQ355618"
2858..3151
/feature="match; GSS: Em:AQ627537.1"
complement(11439..11580)
/feature="match; GSS: Em:AQ031787"
complement(24560..25030)

```

```

/misc_feature /note="match; GSS: Em:AQ307331"
complement(24637..25031)
/misc_feature /note="match; GSS: Em:AQ234890"
27999..28360
/misc_feature /note="match; STS: Em:G14651"
29391..29649
/misc_feature /note="match; STS: Em:G06203"
complement(37609..37922)
/misc_feature /note="match; GSS: Em:AQ114724"
47849..48201
/misc_feature /note="match; GSS: Em:AQ553574.1"
63243..63496
/misc_feature /note="match; GSS: Em:AQ142826"
21006 a 18901 c 18290 g 19125 t
BASE COUNT
ORIGIN

Query Match 15.0%; Score 34.4; DB 10; Length 77322;
Best Local Similarity 63.1%; Pred. No. 1.7;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 14 TGTCTCTCCATCCAGGAGGCGAGTGCCACTATGGGCTCTGGGCTCTGGGCTCTGCTC 73
|||||
Db 31300 TGTCTCTCCATCCAGGAGGCGAGTGCCACTATGGGCTCTGGGCTCTGGGCTCTGCTC 31241
|||||
QY 74 CTCCTGACCTCTCTGGCAGCTCA 97
|||||
Db 31240 CTCCTGACCTCTGGCAGCTCA 31217
|||||

RESULT 5
AC016716/c 21190 bp DNA HTG 04-DEC-1999
LOCUS Homo sapiens clone RP11-31213, *** SEQUENCING IN PROGRESS ***, 31
DEFINITION Unordered pieces.
ACCESSION AC016716
VERSION AC016716.1 GI:6524349
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 21190)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 21190)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 2127: contig of 2127 bp in length
* 2128 2145: gap of unknown length
* 2146 4226: contig of 2081 bp in length
* 4227 4244: gap of unknown length
* 4245 6447: contig of 2203 bp in length
* 6448 6465: gap of unknown length
* 6466 9092: contig of 2627 bp in length
* 9093 9110: gap of unknown length
* 9111 11195: contig of 2083 bp in length
* 11196 11213: gap of unknown length
* 11214 14033: contig of 2820 bp in length
* 14034 14051: gap of unknown length

```


	6
RESULT ACU16764	DNA HTG 09-DEC-1999
LOCUS Homo sapiens clone RP11-549G13, WORKING DRAFT SEQUENCE, 1 unordered pieces.	
DEFINITION ACU16764	
ACCESSION ACU16764.2 GI:6554041	
VERSION ACU16764.2	
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE human.	
ORGANISM Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE 1 (bases 1 to 191010)	
AUTHORS Waterston.R.H.	
TITLE The sequence of Homo saplens clone	
JOURNAL Unpublished	
REFERENCE 2 (bases 1 to 191010)	
AUTHORS Waterston,R.H.	
TITLE Direct Submission	
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA	
COMMENT On Dec 10, 1999 this sequence version replaced gi:6524275. Center project name: _HNE0349G13. * NOTE: This is a "working draft" sequence. It currently * consists of 1 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved * 1 191010: contig of 191010 bp in length. Location/Qualifiers 1..191010 /organism="Homo sapiens" /db_xref="taxon:9606" /cn="RP11-549G13"	
BASE COUNT 46669 A 48592 C 47516 G 47765 T 468 others	
ORIGIN	
Query Match 15.0%; Score 34.4; DB 45; Length 191010; Best Local Similarity 50.6%; Fred.No. 1.7; Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;	
QY 58 CTCGCCOCTTGTCCTCCTTTGCACCCTCTTGGCACGTCAATGCAGCGCCGGGTAT 117	
Db 148271 GCAGAACAATTGCTCTCTCTCGCGACCATCTGGGACTGGCACAAAGCCCAGTGGAG 148330	
QY 118 GACTTTGCAACTGAAGCTGAAGGAGTCTTTCTGTCAAATAATCTCTATGAGTCAGCTT 177	
Db 148331 CCAGGCTTCTCAGAGGGCATGTGACTCACCTGTGTCCCCCTCTCCAACCTTGCAGCTG 148390	
QY 178 CTGTGAATGCTTGAAAAGTCTGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 221	
Db 148391 CAAACCCCTGCAAGAAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 148434	
RESULT 7	
OSU25430	
LOCUS OSU25430 1801 bp mRNA PLN 23-MAY-1995	
DEFINITION Oryza sativa nucleotide pyrophosphatase precursor, mRNA, complete cds	
ACCESSION U25430	
VERSION U25430.1 GI:818848	
KEYWORDS rice.	
SOURCE Osyza sativa	
ORGANISM Eukaryota; Viridiplantae; streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; liliopsida; Poales:	


```

Em:AA673371 Em:AI190893 Em:AI019759 Em:AI770152
Em:AA149582 Em:N30173 Em:AI342072 Em:R54847 Em:AI126154
Em:AA118158 Em:AA044074 Em:N54928 Em:AA215582 Em:AA398600
Em:AI623440 Em:AI024124 Em:AA637284 Em:AI1824779
Em:AI094871 Em:AI192647 Em:N63934 Em:AV083921"
/evidence-not_experimental
/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12171..12948,14414..14546,
14669..14747,15784..15895,15983..16087)
/gene="dJ858B16.1"
/note="match: cDNAs: Em:AB011114"
/evidence-not_experimental
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
273..16218
/gene="dJ858B16.1"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12171..12948,14414..14546,
14669..14747,15784..15895,15983..16087)
/gene="dJ858B16.1"
/note="different from published protein but supported by
GENES and GENSCAN"
/codon_start=1
/evidence-not_experimental
/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"
/protein_id="CAB56392.1"
/db_xref="GI:5921489"
/translation="QAILHAERQLLYRSFWFHHQAAARHQBQEWQTVACAHRRGL
KFAFLWRESAQLRTGRTGRVRAAEFMAQLLWASQWRECLALGAEKQIMRAD
LHQSVLHRLAQAQWTVYQGRVRSILREVAARESOHNRQLLEGALRWKENTWARYDE
AKTQASVHYRTICSKYLVQVREAVSVQNYRQEDCAIWEAKVLDRCGLRTWFO
RWDCSRSAQQLERAVQHRRQLLEGLARWTHHLQCVRRLLHRSQSTQLLAQ
RLSRTCFQWRQQLARROEQRTATRLALFWFASLQAQVWATNLFAVRECKRLMRAQ
WALQYQQLQEGATRLRFAASMAKSRQQLQAQQVAAHSLHRAVRECKRLMRAQ
VIGRHPQKPLAIAFSRKVTEGPLNRIAGAGDGTLETRKQASRPLGALGRLLA
EPFHALELPQKPHGMAQPAAPSLTRPFLAEATPLVPHSPGLSAPGPQKP
PTASTGPELLLLPLSSPCGGAAPARVQSRATPRDKPPVPSLSASVDPHLLPGD
FSNTRAGSLTAGSLDLEALEEIQQLLYHTTKONLWSCQASLSRLRLNRE
EPGPDQVEQVQKEQLEQVEMQIOLLAELQKQROPIGACVARIQLRQALC"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11528)
/gene="dJ858B16.1"
/note="original published form; could be artefactual as it
is result of a large intron read-through; match: proteins:
Tr:O60289"
/codon_start=1
/evidence-not_experimental
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
/protein_id="CAB56393.1"
/db_xref="GI:5921490"
/translation="QAILHAERQLLYRSFWFHHQAAARHQBQEWQTVACAHRRGL
KFAFLWRESAQLRTGRTGRVRAAEFMAQLLWASQWRECLALGAEKQIMRAD
LHQSVLHRLAQAQWTVYQGRVRSILREVAARESOHNRQLLEGALRWKENTWARYDE
AKTQASVHYRTICSKYLVQVREAVSVQNYRQEDCAIWEAKVLDRCGLRTWFO
RWDCSRSAQQLERAVQHRRQLLEGLARWTHHLQCVRRLLHRSQSTQLLAQ
RLSRTCFQWRQQLARROEQRTATRLALFWFASLQAQVWATNLFAVRECKRLMRAQ
WALQYQQLQEGATRLRFAASMAKSRQQLQAQQVAAHSLHRAVRECKRLMRAQ
VIGRHPQKPLAIAFSRKVTEGPLNRIAGAGDGTLETRKQASRPLGALGRLLA
EPFHALEH"
620..913
/note="AluX repeat: matches 1..294 of consensus"
1444..1687
/gene="dJ858B16.1"
/note="match: GSS: Em:AQ545234"
3350..3536
/note="AluJo repeat: matches 2..302 of consensus"
3634..3836
/note="L2 repeat: matches 2497..2701 of consensus"
4356..4646
/note="AluX repeat: matches 21..312 of consensus"

```

```

repeat_region
4650..4754
/note="AluSg/x repeat: matches 200..307 of consensus"
4990..5281
/note="AluX repeat: matches 1..292 of consensus"
5805..6231
/note="MELC repeat: matches 6..460 of consensus"
6529..6668
/note="FRAM repeat: matches 1..140 of consensus"
6671..6966
/note="Alu repeat: matches 1..295 of consensus"
6985..7081
/note="L2R40a repeat: matches 1..100 of consensus"
7929..8097
/note="MIR repeat: matches 82..255 of consensus"
9765..10068
/note="Alu repeat: matches 1..301 of consensus"
10059..10096
/note="14 copies 2 mer ta 100% conserved"
10104..10399
/note="AluJb repeat: matches 1..293 of consensus"
12099..12259
/note="MIR repeat: matches 46..208 of consensus"
13063..13779
/note="L2 repeat: matches 1570..2750 of consensus"
15108..15404
/note="AluSg repeat: matches 1..295 of consensus"
complement(16160)
/gene="dJ858B16.2"
complement(join(16160..17505,18222..18382,18665..18811,
19003..19141,19318..19554,21353..21518,23432..23685,
28473..28494))
/gene="dJ858B16.2"
/note="match: cDNAs: Em:M62722 Em:AL050371 Em:AF086277
Em:AF05304 Em:U82163 Em:AB011114; match: ESTs:
Em:AA324939 Em:AA682112 Em:AA166445 Em:AA166440
Em:AA260704 Em:AI019759 Em:AL039237 Em:AA118158
Em:AV083921 Em:AI232370 Em:A0035550 Em:AV122327
Em:AA717291 Em:AA326672 Em:A0051210 Em:AA298268
Em:AA144290 Em:AA673371 Em:R00603 Em:AA297077 Em:AA237290
Em:AA615772 Em:AA297754 Em:AA297076 Em:AA643609
Em:AI094871 Em:AA298523 Em:AA149582 Em:AA836285
Em:AV031353 Em:AA297159 Em:AI191287 Em:AA916775 Em:R51102
Em:N30173 Em:AA521233 Em:AA044074 Em:AI361548 Em:H69648
Em:AA971234 Em:W00857 Em:AI623440 Em:AA553345 Em:H93993
Em:AA046738 Em:AA637284 Em:AA151798 Em:W69452 Em:AA297086
Em:W69536 Em:T10522 Em:AI024124 Em:H71693 Em:AA302951
Em:N98260 Em:N63934 Em:AA215768 Em:R94341 Em:RA177542
Em:RA8874 Em:AI342072 Em:H90666 Em:AA255719 Em:AI024249
Em:AA829254 Em:W69460 Em:AA595561 Em:W69544 Em:TT7685
Em:AA779744 Em:N64034 Em:AV075750 Em:H38339 Em:AV095355
Em:AV116402 Em:AI425026 Em:AA974403 Em:AA908164
Em:AA088219 Em:AA298551 Em:N56896 Em:AV030271 Em:AI28571
Em:AI806913 Em:AI014094 Em:H53175 Em:AI809747 Em:H83743
Em:H62728 Em:W03262 Em:R52264 Em:AA044185 Em:R82229
Em:AA369042 Em:AA393183 Em:AI126154 Em:R54847 Em:AA158770
Em:AA825239 Em:R53399 Em:AI770152 Em:AA594535 Em:AA582394

```

Query Match 14.88; Score 34; DB 11; Length 36676;

Best Local Similarity 54.99; Pred. No. 2.1;

Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 15 GTCTCTCCATCCAGGAGCCAGTGGCCACTATGGGCTCTGGGCTGCGCCCTTGCTTC 74

Db 14428 GGCTACTCTGAGGATAGCCCCGGTCCCTCATCCCTGGCCAGTGTCCCTGACCCCA 14487

QY 75 TCTTGACCTCTCTGGAGCTACATGGAACAGGCGCGGTATGACTTTGCAACTGAAGC 134

Db 14488 TCTACTCTCTCTGGGAGCTTCTCAGCCACAGGCTGGGCTTTCACCTGACG 14547

QY 135 TG 136

Db 14548 TG 14549

RESULT 9
AC008878
LOCUS Homo sapiens chromosome 19 clone CITB-HL_2207023, 31-OCT-1999
DEFINITION AC008878 291288 bp DNA HTG
PROGRESS ***, 160 unordered pieces.
ACCESSION AC008878
VERSION AC008878.2 GI:6165135
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 291288)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 31, 1999 this sequence version replaced gi:5686195.
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 160 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1016: contig of 1016 bp in length
* gap of unknown length
* 1017 1601: contig of 585 bp in length
* gap of unknown length
* 1602 2427: contig of 826 bp in length
* gap of unknown length
* 2428 3083: contig of 656 bp in length
* gap of unknown length
* 3084 3959: contig of 876 bp in length
* gap of unknown length
* 3960 7980: contig of 3921 bp in length
* gap of unknown length
* 7981 8576: contig of 696 bp in length
* gap of unknown length
* 8577 12400: contig of 3824 bp in length
* gap of unknown length
* 12401 13210: contig of 810 bp in length
* gap of unknown length
* 13211 13929: contig of 719 bp in length
* gap of unknown length
* 13930 16891: contig of 2962 bp in length
* gap of unknown length
* 16892 17522: contig of 631 bp in length
* gap of unknown length
* 17523 18054: contig of 532 bp in length
* gap of unknown length
* 18055 18736: contig of 682 bp in length
* gap of unknown length
* 18737 19619: contig of 883 bp in length
* gap of unknown length
* 19620 20466: contig of 847 bp in length
* gap of unknown length
* 20467 21409: contig of 943 bp in length
* gap of unknown length
* 21410 22300: contig of 891 bp in length
* gap of unknown length
* 22301 23263: contig of 963 bp in length
* gap of unknown length
* 23264 24270: contig of 1007 bp in length
* gap of unknown length
* 24271 25014: contig of 744 bp in length
* gap of unknown length
* 25015 25665: gap of unknown length
* contig of 651 bp in length
* 25666 27090: gap of unknown length
* contig of 1425 bp in length
* 27091 28065: gap of unknown length
* contig of 975 bp in length
* 28066 28837: gap of unknown length
* contig of 772 bp in length
* 28838 32347: gap of unknown length
* contig of 3510 bp in length
* 32348 33118: gap of unknown length
* contig of 771 bp in length
* 33119 33757: gap of unknown length
* contig of 639 bp in length
* 33758 34621: gap of unknown length
* contig of 864 bp in length
* 34622 35157: gap of unknown length
* contig of 536 bp in length
* 35158 36196: gap of unknown length
* contig of 1039 bp in length
* 36197 37333: gap of unknown length
* contig of 1137 bp in length
* 37334 38054: gap of unknown length
* contig of 721 bp in length
* 38055 38762: gap of unknown length
* contig of 708 bp in length
* 38763 39524: gap of unknown length
* contig of 762 bp in length
* 39525 40176: gap of unknown length
* contig of 652 bp in length
* 40177 42851: gap of unknown length
* contig of 2675 bp in length
* 42852 43760: gap of unknown length
* contig of 909 bp in length
* 43761 44291: gap of unknown length
* contig of 531 bp in length
* 44292 44840: gap of unknown length
* contig of 549 bp in length
* 44841 45627: gap of unknown length
* contig of 787 bp in length
* 45628 46298: gap of unknown length
* contig of 671 bp in length
* 46299 47167: gap of unknown length
* contig of 869 bp in length
* 47168 47878: gap of unknown length
* contig of 711 bp in length
* 47879 48711: gap of unknown length
* contig of 833 bp in length
* 48712 48959: gap of unknown length
* contig of 248 bp in length
* 48960 50309: gap of unknown length
* contig of 1350 bp in length
* 50310 51028: gap of unknown length
* contig of 719 bp in length
* 51029 51872: gap of unknown length
* contig of 844 bp in length
* 51873 52523: gap of unknown length
* contig of 651 bp in length
* 52524 53417: gap of unknown length
* contig of 894 bp in length
* 53418 54178: gap of unknown length
* contig of 761 bp in length
* 54179 54928: gap of unknown length
* contig of 750 bp in length
* 54929 55629: gap of unknown length
* contig of 701 bp in length
* 55630 56741: gap of unknown length
* contig of 1112 bp in length
* 56742 58251: gap of unknown length
* contig of 1510 bp in length
* 58252 59103: gap of unknown length
* contig of 852 bp in length
* gap of unknown length

```

* 59104 60607: contig of 1504 bp in length
* gap of unknown length
* 60608 61642: contig of 1035 bp in length
* gap of unknown length
* 61643 62432: contig of 790 bp in length
* gap of unknown length
* 62433 62684: contig of 252 bp in length
* gap of unknown length
* 62685 64028: contig of 1344 bp in length
* gap of unknown length
* 64029 68403: contig of 4375 bp in length
* gap of unknown length
* 68404 69905: contig of 1502 bp in length
* gap of unknown length
* 69906 70653: contig of 748 bp in length
* gap of unknown length
* 70654 72376: contig of 1723 bp in length
* gap of unknown length
* 72377 73174: contig of 798 bp in length
* gap of unknown length
* 73175 73783: contig of 609 bp in length
* gap of unknown length
* 73784 74652: contig of 869 bp in length
* gap of unknown length
* 74653 75356: contig of 704 bp in length
* gap of unknown length
* 75357 76056: contig of 700 bp in length
* gap of unknown length
* 76057 76965: contig of 909 bp in length
* gap of unknown length
* 76966 77981: contig of 1016 bp in length
* gap of unknown length
* 77982 79239: contig of 1258 bp in length
* gap of unknown length
* 79240 80222: contig of 983 bp in length
* gap of unknown length
* 80223 81434: contig of 1212 bp in length
* gap of unknown length
* 81435 82163: contig of 729 bp in length
* gap of unknown length
* 82164 82879: contig of 716 bp in length
* gap of unknown length
* 82880 83520: contig of 641 bp in length
* gap of unknown length
* 83521 83612: contig of 92 bp in length
* gap of unknown length
* 83613 83759: contig of 147 bp in length
* gap of unknown length
* 83760 84424: contig of 665 bp in length
* gap of unknown length
* 84425 85182: contig of 758 bp in length
* gap of unknown length
* 85183 86189: contig of 1007 bp in length
* gap of unknown length
* 86190 87102: contig of 913 bp in length
* gap of unknown length
* 87103 87755: contig of 653 bp in length
* gap of unknown length

Query Match 14.88; Score 34; DB 42; Length 291288;
Best Local Similarity 51.38; Pred. No. 2.4;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 58 GCTGCCCTTGCTCTCTTGTACCCCTTGGCAGCTCAGTGAACGAGCGCGGTAT 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148943 GCTGCTCTTGCCTGCTGCCATCTCTTTTCATTTTATTGAGACCGAGTCT 149002
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 118 GACTTTGCACTGAGCTGAGAGTCTTTTCTGACAAATCTCTATGAGTCCAGCTT 177
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149003 CGCTCTGTATCGAGCTGAGTACAAATGTCGATCTCAGTCACTAATCTCTGCC 149062
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 178 CCTGGAATGCTTTGAAAGTCTGCCCTCTCTCC 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149063 CCTGGTTTCAGTGATTCATGCTCTCTCC 149096
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 10

```

AC016298
LOCUS Homo sapiens clone RP11-655M17, LOW-PASS SEQUENCE SAMPLING. 24-NOV-1999
DEFINITION AC016298
ACCESSION AC016298
VERSION AC016298.1 GI:6467036
KEYWORDS HIG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 62181)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-655M17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 62181)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckgaiter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferrelira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meltrin, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teschaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission,
Submitted (24-NOV-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submission@genome.wi.mit.edu
----- Project Information
Center project name: L4315
Center clone name: 655_M.17
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 775: contig of 775 bp in length
* gap of unknown length
* 776 1583: contig of 808 bp in length
* gap of unknown length
* 1584 2371: contig of 788 bp in length
* gap of unknown length
* 2372 3185: contig of 814 bp in length
* gap of unknown length
* 3186 3980: contig of 795 bp in length
* gap of unknown length
* 3981 4778: contig of 798 bp in length
* gap of unknown length
* 4779 5584: contig of 806 bp in length
* gap of unknown length
* 5585 6392: contig of 808 bp in length

```

TITLE

JOURNAL

COMMENT

35214	36022:	contig of 809 bp in length
		gap of unknown length
36023	36829:	contig of 807 bp in length
		gap of unknown length
36830	37623:	contig of 794 bp in length
		gap of unknown length
37624	38411:	contig of 788 bp in length
		gap of unknown length
38412	39213:	contig of 802 bp in length
		gap of unknown length
39214	40010:	contig of 797 bp in length
		gap of unknown length
40011	40741:	contig of 731 bp in length
		gap of unknown length
40742	41570:	contig of 829 bp in length
		gap of unknown length
41571	42298:	contig of 728 bp in length
		gap of unknown length
42299	42980:	contig of 682 bp in length
		gap of unknown length
42981	43786:	contig of 806 bp in length
		gap of unknown length
43787	44584:	contig of 798 bp in length
		gap of unknown length
44585	45392:	contig of 808 bp in length
		gap of unknown length
45393	46206:	contig of 814 bp in length
		gap of unknown length
46207	46991:	contig of 785 bp in length
		gap of unknown length
46992	47813:	contig of 822 bp in length
		gap of unknown length
47814	48623:	contig of 810 bp in length
		gap of unknown length
48624	49437:	contig of 814 bp in length
		gap of unknown length
49438	50183:	contig of 746 bp in length
		gap of unknown length
50184	50985:	contig of 802 bp in length
		gap of unknown length
50986	51800:	contig of 815 bp in length
		gap of unknown length
51801	52595:	contig of 795 bp in length
		gap of unknown length
52596	53395:	contig of 800 bp in length
		gap of unknown length
53396	54205:	contig of 810 bp in length
		gap of unknown length
54206	54996:	contig of 791 bp in length
		gap of unknown length
54997	55783:	contig of 787 bp in length
		gap of unknown length
55784	56587:	contig of 804 bp in length
		gap of unknown length
56588	57389:	contig of 802 bp in length
		gap of unknown length
57390	58203:	contig of 814 bp in length
		gap of unknown length

	Query Match	14.0%;	Score 33.8;	DB 44;	Length 62181;
	Best Local Similarity	53.4%;	Pred. No. 2.5;		
	Matches 71;	Conservative 0;	Mismatches 62;	Indels 0;	Gaps
QY	6	GACVTCAGTGTCTCTCCATCCAGGAGCGCATGTGCCACATATGGGTCTGTGGGTGCCCC	65		
Db	11905	GATTTTAAATCTCTACCCCTGAAGCTGGAGATGGTCTCTTTGGTTGGGTGCCCC	11964		
QY	66	TTFGCTCTCTTGACCCCTCTTGGCAGCTCATCATGGAACAGGCGCGGTATGACTTTGC	125		
Db	11965	TCCTTCCTCTTGGCCCTGGGTGGCAGCTCCCTGGAAAGAGTCTCATACATGCCCTTGC	12024		
QY	126	AACTGAAGCTGAA	138		

Db 12025 GCGTGGGTGCA 12037

RESULT 11
HSJ773A18/cLOCUS
DEFINITION

HSJ773A18 128379 bp DNA PRI 23-NOV-1999
Human DNA sequence from clone 773A18 on chromosome 1p13.2-21.1
Contains genes for RAS-RELATED PROTEIN RAP-1A (C21KG) (KREV-1
PROTEIN) (GTP-BINDING PROTEIN SMG-P21A) (G-22K), KCND3 (potassium
voltage-gated channel, Shal-related subfamily, member 3), PROBABLE
ATP-DEPENDENT RNA HELICASE P47 HOMOLOG, ESTS, STSS, GSSs and CPG
islands, complete sequence.

ACCESSION

AL049557

VERSION

AL049557.19

KEYWORDS

HTG; C21KG; CPG Island; G-22K; GTP-BINDING PROTEIN; KCND3; KREV-1

SOURCE

human;

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 128379)

TITLE

Coville.G.

JOURNAL

Direct Submission

Submitted (03-SEP-1999)

Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

COMMENT

On Jul 27, 1999 this sequence version replaced g1:5579000.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.: EMBL; Sv.: SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C-elegans/wormpep

the entire insert of clone 773A18. This sequence has been finished

according to sequence map criteria as follows. An attempt is made

to resolve all sequencing problems, such as compressions and

repeats, but not necessarily within known annotated human repeat

sequence elements (e.g. Alu). Where the sequence is ambiguous,

there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of

human chromosome 1, constructed by the Sanger Centre Chromosome 1

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chrl

773A18 is from the library RPCI-4 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see http://bacpac.med.buffalo.edu/VBCTOR: PCYPAC2.

FEATURES

Location/Qualifiers

1..128379

/organism="Homo sapiens"

/db_xref="taxon:9606"

/Chromosome="1"

/map="pl3.2-21.1"

/clone_lib="RPCI-4"

/clone="RP4-773A18"

complement(1..988)

/gene="dj773A18.3"

complement(1..988)

/gene="dj773A18.3"

/note="dj773A18.3 (similar to cytokeratin 18); match;

cDNAs: Em:X12876; match: proteins: Sw:P05784 Sw:P05783"

/codon_start=1

/pseudo

/evidence=not_experimental

1487..1866

misc_feature

1535..1556

repeat_region

1557..1697

repeat_region

/note="MIR repeat: matches 76..206 of consensus"
1872..1977
/note="U6 repeat: matches 2..107 of consensus"
1979..2236
/note="L1MCL1 repeat: matches 6071..6323 of consensus"
2237..2539
/note="AluY repeat: matches 4..299 of consensus"
2540..2597
/note="L1MCL1 repeat: matches 6013..6071 of consensus"
2915..3217
/note="AluX repeat: matches 1..303 of consensus"
3222..3352
/note="AluSc repeat: matches 1..131 of consensus"
3414..3702
/note="AluX repeat: matches 13..301 of consensus"
3855..4028
/note="MIR58C repeat: matches 29..84 of consensus"
4074..4274
/note="L1MB6 repeat: matches 5411..5607 of consensus"
4316..4840
/note="L1MB6 repeat: matches 5594..6136 of consensus"
4841..5150
/note="AluSp repeat: matches 1..309 of consensus"
5151..5192
/note="L1MB6 repeat: matches 6136..6175 of consensus"
5776..5852
/note="MIR repeat: matches 129..209 of consensus"
5862..6106
/note="MIR repeat: matches 5..252 of consensus"
6123..6223
/note="MIR repeat: matches 28..138 of consensus"
6330..6483
/note="FRAM repeat: matches 1..156 of consensus"
6666..6727
/note="31 copies 2 mer aa 77% conserved"
7100..7429
/note="AluY5 repeat: matches 1..308 of consensus"
7838..8216
/note="L2 repeat: matches 1148..1537 of consensus"
8268..8525
/note="L1MB3 repeat: matches 5916..6183 of consensus"
8581..8904
/note="MIR468 repeat: matches 2..236 of consensus"
8906..9254
/note="L1M4 repeat: matches 4791..5146 of consensus"
9234..9548
/note="L1M4 repeat: matches 2658..2976 of consensus"
9550..9605
/note="Tigger3(Golem) repeat: matches 2970..3028 of
consensus"
9606..9702
/note="L1M4 repeat: matches 4324..4430 of consensus"
9730..9809
/note="AluJ/FRAM repeat: matches 228..305 of consensus"
9821..10068
/note="Tigger3(Golem) repeat: matches 2277..2547 of
consensus"
10057..10171
/note="Tigger3(Golem) repeat: matches 1..111 of consensus"
10172..10563
/note="L1MEC repeat: matches 2272..2658 of consensus"
10564..10876
/note="MIR44A repeat: matches 1..333 of consensus"
10877..10971
/note="L1MEC repeat: matches 2480..2272 of consensus"
10972..11047
/note="L1PA2 repeat: matches 6069..6144 of consensus"
11048..11422
/note="L1MEC repeat: matches 2112..2481 of consensus"
11474..11935
/note="L1MEC repeat: matches 1520..2023 of consensus"
11938..12338
/note="L2 repeat: matches 1958..2404 of consensus"

```
repeat_region 12484..12837 /note="L2 repeat: matches 2029..2370 of consensus"
repeat_region 12960..13258 /note="AluSg repeat: matches 1..299 of consensus"
repeat_region 13555..13859 /note="AluY repeat: matches 1..305 of consensus"
repeat_region 14240..14335 /note="MIR repeat: matches 38..140 of consensus"
repeat_region 14350..14652 /note="AluYb8 repeat: matches 1..315 of consensus"
misc_feature complement(14862..15301) /note="match: ESTs: Em:AI673310"
repeat_region 14948..15005 /note="MIR repeat: matches 171..228 of consensus"
repeat_region 15170..15237 /note="L2 repeat: matches 2682..2749 of consensus"
misc_feature <15266..>15532 /note="match: ESTs: Em:AA837720"
repeat_region 15533..15844 /note="AluX repeat: matches 1..312 of consensus"
misc_feature 15942..16366 /note="match: ESTs: Em:AA772504"
repeat_region 16244..16342 /note="MIR repeat: matches 141..262 of consensus"
misc_feature 17326..>17655 /note="match: ESTs: Em:W03890"
repeat_region 18127..18623 /note="L2 repeat: matches 1983..2543 of consensus"
repeat_region 18844..19138 /note="AluX repeat: matches 1..297 of consensus"
misc_feature complement(19286..>19694) /note="match: ESTs: Em:AI281646"
misc_feature 19368..19665 /note="match: ESTs: Em:N76756"
repeat_region 19695..19862 /note="L1Mc repeat: matches 1049..1219 of consensus"
repeat_region 19863..20101 /note="MIR repeat: matches 2..261 of consensus"
repeat_region 20102..20186 /note="L1Mc repeat: matches 1219..1304 of consensus"
repeat_region 20187..20493 /note="AluJb repeat: matches 1..311 of consensus"
repeat_region 20494..20853 /note="L1Mc repeat: matches 1304..1454 of consensus"
repeat_region 20747..21111 /note="MIR repeat: matches 111..516 of consensus"
repeat_region 21166..21597 /note="L1Mc repeat: matches 1509..1991 of consensus"
repeat_region 21633..21941 /note="AluJb repeat: matches 3..308 of consensus"
repeat_region 21945..22194 /note="L1M4 repeat: matches 3044..3318 of consensus"
misc_feature complement(22225..22726) /note="match: GSS: Em:AQ180457"
repeat_region 22604..22705 /note="FRAM repeat: matches -2..120 of consensus"
repeat_region 22908..23214 /note="AluX repeat: matches 1..310 of consensus"
repeat_region 23215..23635

Query Match 14.7%; Score 33.6; DB 11; Length 128379;
Best Local Similarity 48.9%; Pred. No. 3.1;
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 4 GGACATTCAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGGTGCC 63
Db 97429 GTGACCTCTTTTCTCTCACCCTATGGTCTCAGGAGGAGACAGGCGCTTTCTTC 97370

QY 64 CCTGTGCTCTCTTGTGACGCTACATGGAACGCGCGGTATGACTTT 123
Db 97369 CCCTTGTCTATATGCTCTCCCTGGCAGCTGCCACACGTCGCTCTACAGCTA 97310

QY 124 GCACTGAGCTGAAGGAGTCTTTTCTGACAAATTCCTCTATGAGTCCAGCTTCCTGGA 183
```

```
Db 97309 GCTAGTGGAGTGGCCAAAGTGGTATGTGATAAATAGACCTTTGCTCTCAGTTCAGGCC 97250
QY 184 ATTG 187
Db 97249 ATGG 97246

RESULT 12
AC006445 200000 bp DNA HTG 15-SEP-1999
LOCUS Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 7
DEFINITION unorderd pieces.
ACCESSION AC006445
VERSION AC006445.10 GI:5881529
KEYWORDS HTG; HTGS_PHASEI.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 200000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
JOURNAL Direct Submission
TITLE Unpublished
REFERENCE 2 (bases 1 to 200000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
JOURNAL Direct Submission
TITLE Submitted (28-JAN-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT On Sep 14, 1999 this sequence version replaced gi:5822658.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1169: contig of 1169 bp in length
* 1170 5668: gap of unknown length
* 12740: contig of 7072 bp in length
* 12741 17239: gap of unknown length
* 17240 27772: contig of 10533 bp in length
* 27773 32270: gap of unknown length
* 32271 67826: contig of 35556 bp in length
* 67827 72325 108800: contig of 36476 bp in length
* 108801 113298: gap of unknown length
* 113299 160850: contig of 47552 bp in length
* 160851 165348: gap of unknown length
* 165349 200000: contig of 34652 bp in length.
FEATURES
source
1..200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
BASE COUNT 50025 a 35744 c 35457 g 51783 t 26991 others
ORIGIN
```

```
Query Match 14.7%; Score 33.6; DB 41; Length 200000;
Best Local Similarity 57.7%; Pred. No. 3.1;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 108 GCGCGGGTATGACTTTGCACTGAAGTGAAGGAGTCTTTTCTGACAAATTCCTCTATG 167
Db 36776 GACAGGGTATCACTCTGTGTGCCAAGCTAAGTGCAGTTACATGATAATGACTCACTGA 36835

QY 168 AGTCACGCTTCGGAATTCCTTAAAGTCTGCTCCTCCTCC 211
Db 36836 ACCTCAAAATTCGGAATTCCTTCTGCTTGGCTTC 36879
```



```

RESULT 13
AC014817/c
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, 16-NOV-1999
DEFINITION pieces.
ACCESSION AC014817
VERSION AC014817.1 GI:6436518
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 26065)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210440 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..26065
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 7321 a 5960 c 5929 g 6855 t
ORIGIN

Query Match 14.7%; Score 33.6; DB 43; Length 26065;
Best Local Similarity 59.4%; Pred. No. 2.8;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 ACCGGGACTGAGTCTCTCCATCCAGGAGCGAGTGGCCACTATGGGTCCTGGGCT 60
Db 1415 ACCAGGCGCATGATGCTTCTCTCCAGGCGCTCATGGAGCGCTTTTGGCTCCGCT 1356

Qy 61 GCGCTTGTCTCTCTCTGACCCCTCTTGGCAGCTC 96
Db 1355 GCGGCTAAATGCCACCCCTCTCTTGGCAGCTC 1320

RESULT 14
AC017080/c
LOCUS Homo sapiens clone Rp11-469G4, *** SEQUENCING IN PROGRESS ***, 34
DEFINITION unordered pieces.
ACCESSION AC017080
VERSION AC017080.1 GI:6554067
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106791)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 106791)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT Center project name: H_NH0469G04.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2077: contig of 2077 bp in length
2097: gap of unknown length
2098: contig of 4000 bp in length
6098: contig of unknown length
6117: gap of unknown length
6118: contig of 2689 bp in length
8807: gap of unknown length
8827: contig of 2729 bp in length
11555: contig of 2277 bp in length
11576: gap of unknown length
13852: contig of 2277 bp in length
13853: gap of unknown length
13873: contig of 2968 bp in length
16840: gap of unknown length
16861: contig of 2494 bp in length
19354: gap of unknown length
19375: contig of 2827 bp in length
22201: contig of 4000 bp in length
22219: gap of unknown length
26220: contig of 4000 bp in length
26238: gap of unknown length
28239: contig of 2301 bp in length
28540: gap of unknown length
28558: gap of unknown length
31633: contig of 3075 bp in length
31632: gap of unknown length
35652: contig of 4000 bp in length
35671: gap of unknown length
35672: contig of 4000 bp in length
39672: gap of unknown length
39691: contig of 2625 bp in length
42316: gap of unknown length
42334: gap of unknown length
45334: contig of 4000 bp in length
45335: gap of unknown length
48354: contig of 2337 bp in length
48709: gap of unknown length
51147: contig of 2438 bp in length
51166: gap of unknown length
53808: contig of 2642 bp in length
53809: gap of unknown length
55891: contig of 2064 bp in length
55910: gap of unknown length
58044: contig of 2134 bp in length
58063: gap of unknown length
62063: contig of 4000 bp in length
62082: gap of unknown length
66083: contig of 4000 bp in length
66101: gap of unknown length
68112: contig of 2011 bp in length
68132: gap of unknown length
70858: gap of unknown length
70877: contig of 4000 bp in length
74877: gap of unknown length
74895: contig of 4000 bp in length
78896: gap of unknown length
78914: contig of 4000 bp in length
82915: gap of unknown length
82934: gap of unknown length
85818: gap of unknown length
85837: contig of 4000 bp in length
89855: gap of unknown length
93499: contig of 3644 bp in length
93500: gap of unknown length
93519: contig of 4000 bp in length
97519: contig of 4000 bp in length
97538: gap of unknown length
10202: contig of 2664 bp in length
102772: gap of unknown length
102773: contig of 2552 bp in length
102791: gap of unknown length
102792: contig of 4000 bp in length.
FEATURES
Location/Qualifiers
1..106791
/organism="Homo sapiens"

```

```
QAWVTYQGRVRSILRLVEAARESOHNROLLKGALEKRWKENTMARVDEAKKTFAQSTHYH
RTTCSKVLVQWHRRAVSQMYTRQOEDCAIWEAQKVLDRGCLRTWFQRWDCSRRSAQQ
RLQLERAVQHHRHOLLLEGIAWKTHLQCVRKRLLIHQSTQLLAORLSRFTCFRMRRO
QLAARRBORCATVRAALWFASFLOAKVWATLAFVLERRRKARLQWALQAYOGOLLO
EGATRLLREFRASKMKASROLOQAQOOVAOHSHSLHRAVRCATLWKOKVLGGKKPOPLA
AIAPSKVTFEGLPILLNRIANGSGDTLETNRFPQASRPUGALGRALEEPEHALEL"
BASE COUNT      1160 a   1516 c   1601 g   1003 t
ORIGIN

Query Match          14.2%; Score 32.6; DB 9; Length 5280;
Best Local Similarity 54.6%; Pred. No. 5,3;
Matches 65; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY    15  GTCTCTCCATCCCAGGAGCGCAGTGCGCACTATGGGCTCTGGGGCTGCCCTTTGTCTCTCC 74
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    4735 GGCTACTCTCTAGGGGTAAGCCCCCGGTCGCCCTCATCCCTGGCCAGTGTCTCTGACCCCCA 4794

QY    75  TCTTGACCCCTCTCTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTTGCAACTGAAG 133
       ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    4795 TCTACTCTCTCTGGGAGCTTCTCAGCCACCAGGGCTGGGCCTGGGACTTTCAACTGCAG 4853

Search completed: March 21, 2000, 14:42:21
Job time: 3756 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2000, 18:37:52 ; Search time 70.18 seconds
(without alignments)
816.386 Million cell updates/sec

Title: US-09-092-296-2
Perfect score: 229
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCTTCAGGGACCA 229

Scoring table: IDENTITY_NUC
searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	218	95.2	439	1 V84366	Human stomach carc
2	34.4	15.0	2923	1 X04326	Human secreted pro
3	32	14.0	1128	1 T28846	Mouse coxsackievir
4	30.4	13.3	580	1 T02964	Cell-cycle regulat
5	30.4	13.3	1301	1 V50430	Mouse coxsackievir
6	29.8	13.0	3510	1 T18596	RAP-1 radiation pr
7	29.8	13.0	291	1 V86365	EST clone AK34. Ne
8	29	12.7	2259	1 Q05106	Sequence encoding
9	29	12.7	1110	1 N60409	Human apolipoprote
10	29	12.7	1110	1 N50450	Sequence encoding
11	29	12.7	597	1 Q69099	ApoE4L protease CD
12	29	12.7	936	1 Q69101	ApoE4Lx2 protease
13	29	12.7	936	1 T18068	Human ApoE4 cDNA
14	29	12.7	936	1 T18070	Human ApoE4Lx2 cDN
15	29	12.7	1157	1 T06957	Human apolipoprote
16	28.8	12.6	1785	1 T66986	Human brain-specif
17	28.4	12.4	2095	1 Q57953	Human beta1.6-N-ac
18	28.4	12.4	2432	1 T77045	CDNA encoding glyc
19	28	12.2	683	1 Q63794	Bovine trypsinogen
20	28	12.2	701	1 Q63795	Bovine trypsinogen
21	28	12.2	110000	1 V30458_1	Continuation (2 of
22	28	12.2	110000	1 V30459_1	Continuation (2 of
23	27.8	12.1	5359	1 Q50396	Rat WDA receptor
24	27.6	12.1	3088	1 T05828	ADP ribosylation f
25	27.6	12.1	5102	1 V03261	Human COL4A6 gene..
26	27.2	12.0	6030	1 V19007	Rattus norvegicus
27	27.2	11.9	2122	1 T49318	Wnt-10b gene. Expr
28	27.2	11.9	5535	1 X21355	Human BAI1 gene. N
29	27.2	11.9	2161	1 X03034	Human IL-1ra BAC c
30	27.2	11.9	11901	1 Q02998	Human IL-1ra BAC c
31	27	11.8	1554	1 Q71243	Serine protease fo
32	27	11.8	1294	1 T60587	Protein cognate of
33	26.8	11.7	2703	1 Q83695	New Tg-beta famil
34	26.8	11.7	2341	1 T31601	Cartilage-derived
35	26.8	11.7	2703	1 T59729	Human bone morphog
36	26.8	11.7	2703	1 T59405	DNA encoding huma
37	26.8	11.7	2703	1 T61412	DNA encoding huma
38	26.8	11.7	2703	1 T69695	DNA encoding huma
39	26.8	11.7	2703	1 T98191	CDNA for human MP5

40	26.8	11.7	3065	1 V00010	Mouse sex comb on
41	26.8	11.7	3065	1 V00014	Mouse sex comb on
42	26.8	11.7	2703	1 T88340	Human MP52 CDNA. M
43	26.8	11.7	210	1 V19507	Retroviral DNA bas
44	26.8	11.7	116624	1 V52850	Human eyal gene CO
45	26.8	11.7	2090	1 V73869	Rat GLCAT-P CDNA.

ALIGNMENTS

RESULT 1
V84366
ID V84366 standard; cDNA to mRNA; 439 BP.
AC V84366; 1999 (first entry)
DE Human stomach carcinoma cDNA clone HP10408.
KW Transmembrane protein; HP10408; human; stomach cancer; ds.
OS Homo sapiens.
PH Key Location/Qualifiers
FT CDS 75..311
FT /*tag= a
FT /note= "cDNA comprising the coding region (minus
the stop codon) is claimed (Claim 3)"
PN W0985508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI: 99-045730/04.
DR P-PSDB: W88498.
PT New human proteins containing transmembrane domains and their
encoding sequences - useful in the preparation of antibodies and
large-scale protein production, gene diagnosis, and gene therapy
PS Claim 4: Page 135; 178pp; English.
CC This is the nucleotide sequence of cDNA clone HP10408, which
includes a coding region (also claimed) for a novel human
transmembrane protein (see W88498). The clone was isolated from a
stomach cancer cDNA library using a signal sequence detection
method, and by protein synthesis by in vitro translation. The
encoded protein has a putative signal sequence and a putative
internal transmembrane domain. The invention provides nucleotide
sequences (see W84359-76) coding for 18 transmembrane proteins
(see W8491-508), vectors containing such polynucleotides, and
eukaryotic cells containing the vectors. The proteins can be
used as antigens or as compositions in the preparation of
antibodies against the proteins. The polynucleotides can be used
as probes for gene diagnosis, and as gene sources for gene therapy
CC and large-scale production of proteins encoded by the cDNA. The
host cells are used for the detection of ligands corresponding to
the expressed proteins, and the screening of low mol.wt. medicines.
SQ Sequence 439 BP; 89 A; 137 C; 109 G; 104 T;

Query Match	95.2%	Score 218	DB 1	Length 439
Best Local Similarity	99.6%	Pred. No. 2.5e-60		
Matches 229	Conservative 0	Mismatches 0	Indels 1	Gaps 1
QY 1	ACCGGACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGGCT 60			
Db 29	ACCGGACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGGCT 88			
QY 61	GCCTCTGCTCTCTCTGACCTCTCTGCGAGTGCATGGAACAGGCCGGGTATGAC 120			
Db 89	GCCTCTGCTCTCTCTGACCTCTCTGCGAGTGCATGGAACAGGCCGGGTATGAC 148			
QY 121	TTTCAACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTTATGATCCAGCTTCT 180			
Db 149	TTTCAACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTTATGATGATCCAGCTTCT 208			
QY 181	GGAATTCCTTGAAGAG-TCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 229			

PT for diagnosis and treatment of cancer and other diseases involving
 PS abnormal apoptosis
 PS Claim 2; Fig 3; 29pp; English.
 CC A cDNA clone (J18696) codes for a novel radiation-protecting human
 CC checkpoint protein (894906), designated RAP-1, involved in
 CC regulation of cell cycle progression and/or apoptosis. It was
 CC isolated from a human cDNA library established in an immortalised
 CC xeroderma pigmentosa cell line, GM2096 (XP1MI). A cDNA (cDNA3) contg.
 CC a 462 bp insert complemented UV sensitivity and was used to screen a
 CC lambda-gt10-K562 human cDNA library. The insert represented part of
 CC the 3510 bp RAP-1 mRNA. The human checkpoint gene can be used as a
 CC marker for early detection and prevention of tumour progression or
 CC can be used to design new drugs that enhance radiotherapy or
 CC chemotherapy of diseases involving abnormal apoptosis.
 CC Sequence 3510 BP; 374 A; 795 C; 848 G; 892 T;
 SQ Sequence 3510 BP; 374 A; 795 C; 848 G; 892 T;

Query Match 13.0%; Score 29.8; DB 1; Length 3510;
 Best Local Similarity 51.9%; Pred. No. 3.6;
 Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 11 CAGTGTCTCCCTCCATCCAGGAGCGAGTGGCTCTGGGCTGGCCCTTGTC 70
 DB 1771 CACCTGCCCTCCATCCAGGAGCGAGAGAGAGATCACTCTATCTCTCTCTGGA 1830

QY 71 CTCCTCTTGACCTCTCTGGAGCTCAGATGGAACAGGCGCGGTATGACTTTGCACTG 130
 DB 1831 TACCTCTCTGGACTTCTCCAAAGAAACAAAGAAAGAGAGAGATCTAGTTGCCAGCTT 1890

QY 131 AAGCTGAAG 139
 DB 1891 AAACGGAGG 1899

RESULT 7
 V86365/c
 ID V86365 standard; cDNA; 291 BP.
 AC V86365;
 DT 27-APR-1999 (first entry)
 DE EST clone AR34.
 KW Expressed sequence tag; secreted protein; hematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS Homo sapiens.
 FN WO9845435-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06954.
 PR 10-APR-1997; US-835913.
 PA (GENY) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racine LA, Spaulding V, Treacy M;
 DR WPI; 99-070076/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 210; 633pp; English.
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC hematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC therapy. The EST sequences are also stated to be useful for gene
 SQ Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;

Query Match 13.0%; Score 29.8; DB 1; Length 291;
 Best Local Similarity 70.2%; Pred. No. 1.5;
 Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 115 TATGACTTTGCACTGAGCTGAGGAGCTTTTCTGACAAATTCCTCTATGAGTC 171
 DB 257 TTTGATTTGTAACAGAGCTGAAGTATTTTCTTACACATTTTCTCTTAAAAATC 201

RESULT 8
 Q05106
 ID Q05106 standard; DNA; 2259 BP.
 AC Q05106;
 DT 02-NOV-1990 (first entry)
 DE Sequence encoding tumour-associated antigen GA733-1.
 KW Pancreatic carcinoma; GA733-1; cancer; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 307..1275
 FT /*tag= a

PN EP-376746-A.
 PD 4-JUL-1990.
 PF 29-DEC-1989; 313687.
 PR 29-DEC-1988; US-291583.
 PA (WIST-) Wistar Inst.
 PI Linenbach A;
 DR WPI; 90-203091/27.
 DR P-PSDB; R05711.
 PT Tumor-associated antigen, GA733-2 -
 PT Expressed in pancreatic carcinoma cells, used for tumor
 PT immuno-therapy.
 PS Claim 2; Fig 2A; 12pp; English.
 CC The peptide is produced by tumor cells, and Abs raised to the
 CC peptide may be used in diagnosis and therapy of human tumours.
 SQ Sequence 2259 BP; 460 A; 648 C; 639 G; 512 T;

Query Match 12.7%; Score 29; DB 1; Length 2259;
 Best Local Similarity 52.0%; Pred. No. 5.5;
 Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 17 CTCCTCATCCAGGAGCGAGTGGCCACTATGAGGTCTGGGTGCCCTTGTCTCTC 76
 DB 1120 CGCTCATCCCGCGGCTCATCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1179

QY 77 TTGACCTCTCTTGGCAGCTCACATGGAACAGGCGCGGTATGACTTTGCACTGAAGCTG 136
 DB 1180 ATGCCCTCTCTGCTGATCACCACCGAGAGAGTCCGGGAGTACAGAGGTGGAGATC 1239

QY 137 AAGGA 141
 DB 1240 AAGGA 1244

RESULT 9
 N60409
 ID N60409 standard; DNA; 1110 BP.
 AC N60409;
 DT 01-JAN-1980 (first entry)
 DE Human apolipoprotein-E.
 KW Apolipoprotein-E; hyperlipidemia; arteriosclerosis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 15..968
 FT /*tag= a

PN AU8547513-A.
 PD 24-APR-1986.
 PF 17-SEP-1985; 047513.
 PR 11-JUN-1985; JP-126989.
 PA (MITU) MITSUBISHI CHEM IND KK.
 DR WPI; 86-150217/24.
 DR P-PSDB; F60507.

PT related diseases
PS Claim 13; Page 43; 72pp: English.
CC The cDNA encodes APOE4x2, a protease catalyzing the formation of the
CC abnormal beta/A4 variant of beta-amyloid protein, which is used
CC to develop an inhibitor for the diagnosis and treatment of Alzheimer
CC disease, Down syndrome, Parkinson disease, schizophrenia, a
CC hyperlipoproteinemia or cardiovascular disease.
SQ Sequence 936 BF; 149 A; 317 C; 336 G; 134 T;

Query Match 12.7%; Score 29; DB 1; Length 936;
 Best Local Similarity 54.1%; Pred. No. 4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCATATGGGCTCTGGCTGCCCC 65
 DB 305 GAGCGCGGCTCAGCGCCATCCCGAGCGCTTGGGGCCCTGGTGAACAGGCGCGGTG 246

OY 66 TTGTCTCTCTCTTGACCTCTTTGGCAGCTCACATGAACAGGCGCGG 114
 DB 245 CGGCGCGCCTCAGCGCCATCCCGAGCGCTTGGGGCCCTGGTGAACAGGCGCGGTG 246

RESULT 13

T18068/c
 ID T18068 standard; cDNA; 597 BP.
 AC T18068;
 DT 22-MAY-1996 (first entry)
 DE Human ApoE4 cDNA.
 KW Alzheimer disease; beta-amyloid peptide precursor; APP;
 KW antisense; hyperlipoproteinaemia; cardiovascular disease;
 KW gene therapy; apolipoprotein; ApoE4; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds l..597
 FT /*tag= a
 FT /transl_except= pos:505..507;aa:Ala
 FT /transl_except= pos:553..555;aa:Ser

PN CA3126787-A.
 PD 28-DEC-1995.
 PF 27-JUN-1994; 126787.
 PR 27-JUN-1994; CA-126787.
 PA (FRED/) PREDDIE R E.
 PI Bergmann JE, Preddie RE;
 DR WPI; 96-151776/16.
 DR P-PSDB; R92113.
 PT Proteins and nucleic acids associated with Alzheimer's disease -
 PT used to develop products for diagnosis of Alzheimer's disease and
 PT related conditions, hyperlipoproteinaemia or cardiovascular disease
 PS Claim 9: Fig 3A; 67pp; English.
 CC 2 Open reading frames (T18068 and T18069) identified in the
 CC apolipoprotein E (ApoE) antisense sequence code for proteins
 CC apoE4L (R92113) and apoE4L1 (R92114), respectively. Both ORFs
 CC are transcribed from an upstream regulatory region (T18074).
 CC ApoE4 comprises 2 alleles, one of which causes E4L and E4L1 to
 CC combine to form a hybrid ORF, ApoE4Lx2 (T18070). Elucidation of
 CC the significance of these and other sequences (see also T18065-77)
 CC in the aetiology of Alzheimer's disease (AD) provides means for
 CC diagnosing AD and related diseases, for the design of therapeutic
 CC reagents (e.g. ribozymes or antibodies) and potentially for gene
 CC therapy.
 SQ Sequence 597 BP; 63 A; 241 C; 196 G; 97 T;

Query Match 12.7%; Score 29; DB 1; Length 597;
 Best Local Similarity 54.1%; Pred. No. 3.4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCATATGGGCTCTGGCTGCCCC 65
 DB 305 GAGCGCGGCTCAGCGCCATCCCGAGCGCTTGGGGCCCTGGTGAACAGGCGCGGTG 246

OY 66 TTGTCTCTCTCTTGACCTCTTTGGCAGCTCACATGAACAGGCGCGG 114
 DB 245 CGGCGCGCCTCAGCGCCATCCCGAGCGCTTGGGGCCCTGGTGAACAGGCGCGGTG 246

RESULT 14

T18070/c
 ID T18070 standard; cDNA; 936 BP.
 AC T18070;
 DT 22-MAY-1996 (first entry).

DE Human ApoE4Lx2 cDNA.
 KW Alzheimer disease; beta-amyloid peptide precursor; APP;
 KW antisense; hyperlipoproteinaemia; cardiovascular disease;
 KW gene therapy; apolipoprotein; ApoE4Lx2; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds l..597
 FT /*tag= a
 FT /transl_except= pos:505..507;aa:Ala
 FT /transl_except= pos:553..555;aa:Ser

PN CA3126787-A.
 PD 28-DEC-1995.
 PF 27-JUN-1994; 126787.
 PR 27-JUN-1994; CA-126787.
 PA (FRED/) PREDDIE R E.
 PI Bergmann JE, Preddie RE;
 DR WPI; 96-151776/16.
 DR P-PSDB; R92113.
 PT Proteins and nucleic acids associated with Alzheimer's disease -
 PT used to develop products for diagnosis of Alzheimer's disease and
 PT related conditions, hyperlipoproteinaemia or cardiovascular disease
 PS Claim 13: Fig 6A; 67pp; English.
 CC 2 Open reading frames (T18068 and T18069) identified in the
 CC apolipoprotein E (ApoE) antisense sequence code for proteins
 CC apoE4L (R92113) and apoE4L1 (R92114), respectively. Both ORFs
 CC are transcribed from an upstream regulatory region (T18074).
 CC ApoE4 comprises 2 alleles, one of which causes E4L and E4L1 to
 CC combine to form a hybrid ORF, ApoE4Lx2 (T18070). Elucidation of
 CC the significance of these and other sequences (see also T18065-77)
 CC in the aetiology of Alzheimer's disease (AD) provides means for
 CC diagnosing AD and related diseases, for the design of therapeutic
 CC reagents (e.g. ribozymes or antibodies) and potentially for gene
 CC therapy.
 SQ Sequence 936 BP; 149 A; 317 C; 336 G; 134 T;

Query Match 12.7%; Score 29; DB 1; Length 936;
 Best Local Similarity 54.1%; Pred. No. 4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCATATGGGCTCTGGCTGCCCC 65
 DB 305 GAGCGCGGCTCAGCGCCATCCCGAGCGCTTGGGGCCCTGGTGAACAGGCGCGGTG 246

OY 66 TTGTCTCTCTCTTGACCTCTTTGGCAGCTCACATGAACAGGCGCGG 114
 DB 245 CGGCGCGCCTCAGCGCCATCCCGAGCGCTTGGGGCCCTGGTGAACAGGCGCGGTG 246

RESULT 15

T06957
 ID T06957 standard; cDNA to mRNA; 1157 BP.
 AC T06957;
 DT 19-JUN-1996 (first entry)
 DE Human apolipoprotein-E (ApoE) cDNA.
 KW Recombinant; human; apolipoprotein-E; ApoE; insect cells; larva;
 KW Manduca sexta; Autographa californica nuclear polyhedrosis virus;
 KW haemolymph; lipid complex; biologically active; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_feature 62..119
 FT /*tag= a
 FT note= "misc_signal"
 FT 62..1015
 FT /*tag= b

PN US5472858-A.
 PD 05-DEC-1995.
 PF 04-JUN-1991; 709949.
 PR 04-JUN-1991; US-709949.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Attie AD, Beckage NE, Gretsch DG, Sturley SL;
 DR WPI; 96-029812/03.
 DR P-PSDB; R86791.

5-6, 11 14
19-21
27-29 34 30-33
1-4, 12, 22-25
1-3, 1

10-18

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
C 1	30.4	13.3	580	4	PCR-US95-610-7	Sequence 7, Appli	
C 2	30.4	13.3	580	6	PCR-US95-04636-7	Sequence 7, Appli	
C 3	29.8	13.0	3509	3	US-08-817-4368-1	Sequence 1, Appli	
C 4	29.6	12.9	7218	1	US-08-232-463-14	Sequence 14, Appli	
C 5	29	12.7	1157	1	US-07-709-949-1	Sequence 1, Appli	
C 6	29	12.7	2099	7	5185254-3	Patent No. 5185254	
C 7	28.4	12.4	2095	2	US-08-405-230-8	Sequence 8, Appli	
C 8	28.4	12.4	2095	3	US-08-910-990-8	Sequence 8, Appli	
C 9	27.6	12.1	5102	2	US-08-494-168-1	Sequence 1, Appli	
C 10	27.6	12.1	3088	2	US-08-418-4448-1	Sequence 1, Appli	
C 11	27.2	11.9	2122	2	US-08-485-449-1	Sequence 1, Appli	
C 12	27.2	11.9	3398	6	PCR-US95-08493-12	Sequence 12, Appli	
C 13	27	11.8	1554	2	US-08-469-486-1	Sequence 1, Appli	
C 14	27	11.8	1554	4	US-08-469-558-1	Sequence 1, Appli	
C 15	27	11.8	1294	4	US-08-665-647-4	Sequence 4, Appli	
C 16	26.8	11.7	3065	4	US-08-852-153-7	Sequence 7, Appli	
C 17	26.8	11.7	2703	4	US-08-288-5080-1	Sequence 4, Appli	
C 18	26.6	11.6	1980	4	US-08-766-8584-4	Sequence 4, Appli	
C 19	26.6	11.6	1300	4	US-09-166-703-41	Sequence 41, Appli	
C 20	26.4	11.5	3795	1	US-08-375-709-1	Sequence 1, Appli	
C 21	26.4	11.5	8268	1	US-08-375-709-10	Sequence 10, Appli	
C 22	26.4	11.5	37895	2	US-08-752-929-1	Sequence 1, Appli	
C 23	26.4	11.5	8268	2	US-08-752-929-10	Sequence 10, Appli	
C 24	26.4	11.5	1593	3	US-08-878-989-9	Sequence 9, Appli	
C 25	26	11.4	7326	2	US-08-194-468-1	Sequence 1, Appli	
C 26	26	11.4	44377	3	US-08-804-2270-7	Sequence 7, Appli	
C 27	26	11.4	44377	4	US-08-804-198-1	Sequence 1, Appli	
C 28	25.8	11.3	2814	1	US-07-781-034-1	Sequence 1, Appli	
C 29	25.8	11.3	2334	1	US-08-406-070-1	Sequence 2, Appli	
C 30	25.8	11.3	655	3	US-08-812-845-2	Sequence 1, Appli	
C 31	25.8	11.3	2814	6	PCR-US92-08328-1	Sequence 1, Appli	
C 32	25.6	11.2	1944	1	US-08-565-617-1	Sequence 1, Appli	

QY 95 TCACATGGAACGGCGGGTATGACTTTCGCACTG 130
Db 68 AGAAATGGCTCTTCGCCGCGGTGAGATTGCTACAG 33

RESULT 2

PCT-US95-04636-7/c
; Sequence 7, Application PC/TUS9504636
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
; NUMBER OF SEQUENCES: 10
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04636
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/345,147
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,511
; FILING DATE: 14-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,812
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,371
; FILING DATE: 14-APR-1994
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91...480
PCT-US95-04636-7

Query Match 13.3%; Score 30.4; DB 6; Length 580;
Best Local Similarity 57.3%; Pred. No. 0.25;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 35 GCAGTGGCCACTAAGGGTCTGGGCTGCCCTGTGCTCTTCCTGTGACCTCTCTGGCAGC 94
Db 128 GCGGTGGCCAGCGCGCTGCTGCTGCGCCCAACATGCCCTGTGCCCGGCTGTGGC 69
QY 95 TCACATGGAACGGCGGGTATGACTTTCGCACTG 130
Db 68 AGAAATGGCTCTTCGCCGCGGTGAGATTGCTACAG 33

RESULT 3

US-08-817-436A-1
; Sequence 1, Application US/08817436A
; Patent No. 588280
; GENERAL INFORMATION:
; APPLICANT: Canaan, Dan
; TITLE OF INVENTION: Human Checkpoint Gene and Gene for
; TITLE OF INVENTION: Antisense RNA thereof
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wignan, Cohen, Leitner & Myers
; STREET: 900 17th Street, N.W., Suite 1000
; CITY: Washington

; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.5 inch, 1.44mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,436A
; FILING DATE: 11-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12445
; FILING DATE: 11-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cohen, Herbert
; REGISTRATION NUMBER: 25,109
; REFERENCE/DOCKET NUMBER: 0744.066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-463-7700
; TELEFAX: 202-463-6915
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3509 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: ESTABLISHED XERODERMA PIGMENTOSUM GM2096-SV3
; IMMEDIATE SOURCE:
; CLONE: RAP-1 cDNA
US-08-817-436A-1

Query Match 13.0%; Score 29.8; DB 3; Length 3509;
Best Local Similarity 51.9%; Pred. No. 0.91; Mismatches 62; Indels 0; Gaps 0;
Matches 67; Conservative 0;

QY 11 CAGTGTCTCTCCATCCAGAGCGCAGTATGGGCTCTGGGCTGCCCTTTGTC 70
Db 1771 CACGTCCTCCATCCAGAGCGGAGAGATACATCTCTATCTCTCTCTTGA 1830
QY 71 CTCTCTTGACCTCTTGGCAGCTCACATGGAACAGGCGCGGTATGACTTTCGCACTG 130
Db 1831 TACTCTTGACCTCTTCCAAAGAAACAGAGAGAGATCTAGTTGGCAGCTT 1890
QY 131 AAGCTGAAG 139
Db 1891 AAACGGAGG 1899

RESULT 4

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
;
US-08-232-463-14

Query Match 12.9%; Score 29.6; DB 1; Length 7218;
Best Local Similarity 2.3%; Pred. No. 1.5;
Matches 5; Conservative 129; Mismatches 88; Indels 0; Gaps 0;

QY 8 CTTGAGTGTCTCTCCATCCAGGAGCGAGTGGCCACTATGGGTCTGGCTCCGCCCTT 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1219 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1278

OY 68 GTCTCTCTCTGACCCCTCTGCGAGTCACATGGAACAGGCGCGGTATGACTTGCAA 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1279 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1338

QY 128 CTGAAGCTGAGAGAGTCTTTTCACAAATCTCTATGAGTCCAGCTCTCTGGAATTG 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1339 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1398

QY 188 CTGAAAGTCTGCTCTCTCTCCATCCCTTCAGGACCA 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1399 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440

RESULT 5
US-07-709-949-1
; Sequence 1, Application US/07709949
; Patent No. 5472858
; GENERAL INFORMATION:
; APPLICANT: Attie, Alan D.
; APPLICANT: Gretch, Daniel G
; APPLICANT: Sturley, Stephen L
; APPLICANT: Beckage, Nancy E
; TITLE OF INVENTION: Production of Recombinant Proteins in
; TITLE OF INVENTION: Insect Larvae
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/709,949
; FILING DATE: 19910604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 9629691801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1157 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 62..1015
; FEATURE:
; NAME/KEY: misc.signal
; LOCATION: 62..119
; PUBLICATION INFORMATION:
; AUTHORS: McLean, J W
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 259
; PAGES: 6498-6504
; DATE: 1984
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1157
;
US-07-709-949-1

Query Match 12.7%; Score 29; DB 1; Length 1157;
Best Local Similarity 54.1%; Pred. No. 1;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCACTATGGGTCTGGGCTGCCCC 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 626 GAGCGCGGCTCTAGCGCCATCCGCGGCGCTGGGCGCCCTGGTGGAAACAGGCGCGTG 685

QY 66 TTGTCTCTCTCTTGGACCTCTTGGAGCTCATCATGGAACAGGCGCGGG 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 686 CGGGCGGCGCACTGTGGGCTCTCTGCGCGGCGCGCTACAGGAGCGGG 734

RESULT 6
5185254-3
; Patent No. 5185254
; APPLICANT: LINNENBACH, ALBAN
; TITLE OF INVENTION: GENE FAMILY OF TUMOR-ASSOCIATED ANTIGENS
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/291,583
; FILING DATE: 29-DEC-1988
; SEQ ID NO: 3;
; LENGTH: 2259
;
5185254-3

Query Match 12.7%; Score 29; DB 7; Length 2259;
Best Local Similarity 52.0%; Pred. No. 1.4;
```


Matches 80; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 63 CCTGTGCTCTCTTACCTCTTGGAGCTCAGATGGAACAGGCGGCGGTATGACTT 122
DB 487 CCATTGACAAAGATTGTCCACCTTCGACTCCCAACTTGCCTATGCGCTGCAAAATGTTA 428
QY 123 TGAACCTGAAGCTGAAGGAGCTCTTCTGACAAATTCCTCCATGATGATGCTTCTCTG 182
DB 427 TCTAATACACAGCAGGTTTCTTCAGATCATATGATGCTGATGCTCCAGCGTATGGC 368
QY 183 AATTGCTGAAGCTGCTGCT 228
DB 367 CCATCCACCATTCCT 322

RESULT 9

US-08-494-168-1/c
; Sequence 1, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TOPOLOGY: linear
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2..82, 86..97, 101..4399, 4403..4420, 4424
; LOCATION: ..4465, 4469..4876, 4880..5101)
US-08-494-168-1

Query Match 12.1%; Score 27.6; DB 2; Length 5102;

Best Local Similarity 56.7%; Pred. No. 6;

Matches 51; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 2 CCGGACTTCAGTCT 61
DB 3992 CAGGGTCTCCGGTGTGCGCTTGTTCAGATGATGCGCCAGGGGGGCTGGGCTCCACGG 3933
QY 62 CCCCTGTGCT 91

DB 3932 GGCCTGGGCGGCT 3903

RESULT 10

US-08-418-444A-1/c
; Sequence 1, Application US/08418444A
; Patent No. 5773688
; GENERAL INFORMATION:
; APPLICANT: KURODA, HISAO
; APPLICANT: HIROTA, NAOTAKA
; APPLICANT: ITO, KAZUTOSHI
; TITLE OF INVENTION: GENE EXPRESSION VECTOR USING THE GENE
; TITLE OF INVENTION: EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,444A
; FILING DATE: 07-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP HEI 6-71048
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5773688man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-024-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELE: 24855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3088 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-418-444A-1

Query Match 12.1%; Score 27.6; DB 2; Length 3088;

Best Local Similarity 72.0%; Pred. No. 4.8;

Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 41 GCCACTATGGGGTCTGGGCTGCCCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 90
DB 64 GCCAGCAGGCGCTGGGCGAGCCCTTGGCCCCCTCTGGCCCATCTCGG 15

RESULT 11

US-08-485-449-1
; Sequence 1, Application US/08485449
; Patent No. 5824789
; GENERAL INFORMATION:
; APPLICANT: VANDENBERG, DAVID
; TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

Search completed: March 22, 2000, 01:22:33
Job time: 5637 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2000, 05:09:43 ; Search time 802.03 Seconds
(without alignments)
1078.048 Million cell updates/sec

Title: US-09-092-296-2
Perfect score: 229
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCTTCAGGGACCA 229

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST.*

Word size : 0

Number of hits that pass the threshold : 9077368

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*
- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*

- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*
- 54: em_est22.*
- 55: em_est23.*
- 56: em_est24.*
- 57: em_est25.*
- 58: em_est26.*
- 59: gb_est33.*
- 60: gb_est34.*
- 61: gb_est35.*
- 62: gb_est36.*
- 63: gb_est37.*
- 64: gb_est38.*
- 65: em_est27.*
- 66: em_est28.*
- 67: em_est29.*
- 68: em_est30.*
- 69: gb_est39.*
- 70: gb_est40.*
- 71: gb_est41.*
- 72: gb_est42.*
- 73: gb_est43.*
- 74: gb_est44.*
- 75: em_est31.*
- 76: em_est32.*
- 77: em_est33.*
- 78: em_est34.*
- 79: gb_gss1.*
- 80: gb_gss2.*
- 81: gb_gss3.*
- 82: gb_gss4.*
- 83: em_gss1.*
- 84: em_gss2.*
- 85: em_gss3.*
- 86: em_gss4.*
- 87: gb_gss5.*
- 88: gb_gss6.*
- 89: gb_gss7.*
- 90: gb_gss8.*
- 91: gb_gss9.*
- 92: em_gss5.*
- 93: em_gss6.*
- 94: em_gss7.*
- 95: em_gss8.*
- 96: em_gss9.*
- 97: em_gss10.*
- 98: em_gss11.*
- 99: gb_gss10.*
- 100: gb_gss11.*
- 101: em_gss12.*
- 102: gb_gss12.*
- 103: gb_gss13.*
- 104: gb_gss14.*
- 105: gb_gss15.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	113	49.3	404	61	AI857998	AI857998 wj69b01.x
C 2	75.4	32.9	552	82	AQ718761	AQ718761 HS_5511.B
C 3	61.8	27.0	328	42	AI136523	AI136523 UI-R-C2p-
C 4	35	15.3	348	21	F06958	F06958 HSC1QC101.D
C 5	35	15.3	509	88	AQ838514	AQ838514 HS_5011.A
C 6	34.4	15.0	238	32	AA376266	AA376266 EST88915
C 7	34.4	15.0	412	38	AA769782	AA769782 an71b05.s

```

c 8 34.4 15.0 458 49 A1632159 ts85e12.x
c 9 34.4 15.0 470 60 A1803529 tc42h11.x
c 10 32.4 14.1 537 64 AW078074 fe24b05.y
c 11 32.2 14.1 523 69 AW128453 fel6b05.y
c 12 32.2 14.1 339 21 F08745 HSC1DB011.n
c 13 32.2 14.1 742 27 F78408 HSE78408.Hu
c 14 32.2 14.1 477 45 A1351147 gt11g01.x
c 15 32 14.0 360 63 AW009894 ws88h09.x
c 16 31.8 13.9 313 61 A1810515 wb99f06.x
c 17 31.8 13.9 568 100 AQ290553 nbxb0037C
c 18 31.6 13.8 573 23 T41535
c 19 31.6 13.8 438 38 AA797411 vw27f04.r
c 20 31.6 13.8 318 74 AW218386 EST303569
c 21 31.4 13.7 505 20 M89011
c 22 31.2 13.6 883 79 C8S00CVS
c 23 31.2 13.6 859 79 C8S00DED
c 24 31.2 13.6 536 23 T41525
c 25 31 13.5 427 59 A1776288
c 26 31 13.5 368 23 H28604
c 27 30.8 13.4 589 51 A1730494
c 28 30.8 13.4 513 60 A1807980
c 29 30.6 13.4 527 45 A1360929
c 30 30.6 13.4 308 45 A1394138
c 31 30.6 13.4 667 48 A1533740
c 32 30.6 13.4 709 61 A1866474
c 33 30.6 13.4 457 103 AQ500124
c 34 30.6 13.4 433 23 R90433
c 35 30.4 13.3 545 23 T41524
c 36 30.4 13.3 523 23 T44883
c 37 30.4 13.3 548 23 T76643
c 38 30.4 13.3 392 26 W70374
c 39 30.4 13.3 290 33 AA387480
c 40 30.4 13.3 529 37 H76456
c 41 30.4 13.3 497 37 T44613
c 42 30.4 13.3 407 74 AV395857
c 43 30.4 13.3 445 87 AQ768084
c 44 30.4 13.3 535 105 AQ625402
c 45 30.4 13.3

```

ALIGNMENTS

```

RESULT 1
A1857998/c
LOCUS wj69b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408041 3',
DEFINITION mRNA sequence.
ACCESSION A1857998
VERSION A1857998.1 GI:5511614
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llni.gov/bbrp/image/image.html

```

```

Seq primer: -400p from Gibco
High quality sequence stop: 395.
FEATURES
source
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2408041"
/lab_host="NCI_CGAP_Lu19"
/dev_stage="adult"
/note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 97 a 105 c 117 g 84 t 1 others
ORIGIN
Query Match 49.3%; Score 113; DB 61; Length 404;
Best Local Similarity 98.4%; Pred. No. 2.5e-23;
Matches 124; Conservative 0; Mismatches 1; Gaps 1;
QY 105 CAGGCGCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTCT 164
DB 311 CAGGCGCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTCT 252
QY 165 ATGAGTCAGGCTTCCTGGAATTCCTGAAAG-TCGCGCTCCTCCTCCATCCCTCAG 223
DB 251 ATGAGTCAGGCTTCCTGGAATTCCTGAAAGCTGCTGAAAGCTGCTCCTCCATCCCTCAG 192
QY 224 GGACCA 229
DB 191 GGACCA 186
RESULT 2
A1857998/c
LOCUS A1857998 552 bp DNA GSS 13-JUL-1999
DEFINITION HS_5511_B2_F09_J7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1087 Col=18 Row=L, genomic survey sequence.
ACCESSION A1857998
VERSION A1857998.1 GI:5468077
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 95 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from

```

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Resear h Genetics (inforesgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1087 row: L column: 18
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 552.

Location/Qualifiers
 1..552

/db_xref="taxon:9606"

/clone="Plate-1087 Col-18 Row-L"

/clone_lib="RPC1-11 Human Male BAC Library"

/sex="male"

BASE COUNT 141 a 153 c 124 g 117 t 17 others
 ORIGIN
 /note="Vector: pBAC3.6; Genomic sequence of BAC ends"

Query Match 32.9%; Score 75.4; DB 82; Length 552;
 Best Local Similarity 87.2%; Pred. No. 3.4e-12;
 Matches 82; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 103 AACAGGCGGGTATGCTTTCAGCTGAAGTGAAGGCTCTTTCGACAAATCCCTC 162
 Db 474 AGCATGGCGAGGTATGCTTTCAGCTGAAGGCTCTTTCGACAAATCCCTC 415
 QY 163 CTATGAGTCCAGCTTCTCGAATTCCTGAAAG 196
 Db 414 TTATGAGTCCAGCTTCTCGAATTCCTGAAAG 381

RESULT 3
 LOCUS A1136523/c 328 bp mRNA EST 11-FEB-1999
 DEFINITION UI-R-C2p-ng-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
 UI-R-C2p-ng-e-02-0-UI 3', mRNA sequence.
 ACCESSION A1136523
 VERSION A1136523.1 GI:3637300
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus.

REFERENCE 1. (bases 1 to 328)
 Eutheria; Rodentia; Sclurognathi; Muridae; Rattus.
 Bonaudo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 9704477

On Jan 19, 1998 this sequence version replaced gi:2150222.
 CONTACT Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.wesg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult lung library. cDNA Library Preparation: M. Fatima Bonaudo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics

Seq primer: M13 Forward.
 Location/Qualifiers
 1..328
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C2p-ng-e-02-0-UI"
 /clone_lib="UI-R-C2p"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"

FEATURES
 source

/note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-C2p
 library is a subtracted library derived from the UI-R-C1
 library, which is a subtracted library derived from the
 UI-R-C0 library. The UI-R-C0 library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dT track which allows
 identification of the library of origin of a clone within
 the mixture. The subtracted library (UI-R-C2p) was
 constructed as follows: PCR amplified cDNA inserts from
 UI-R-C1 clones from which 3' ESTs had been derived was
 used as a driver in a hybridization with the UI-R-C1
 library in the form of single-stranded circles. The
 remaining single-stranded circles (subtracted library) was
 purified by hydroxyapatite column chromatography,
 converted to double-stranded circles and electroporated
 into DH10B bacteria (Life Technologies) to generate the
 UI-R-C2p library. This procedure has been previously
 described (Bonaudo, Lennon and Soares, Genome Research 6:
 791-806, 1996)

BASE COUNT 62 a 77 c 98 g 91 t
 ORIGIN
 Query Match 27.0%; Score 61.8; DB 42; Length 328;
 Best Local Similarity 69.5%; Pred. No. 2.8e-08;
 Matches 98; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 90 GCAGCTCATGTGAACAGCGCGGGGTATGACTTTCGAACTGAAGCTGCTTTTC 149
 Db 327 GCAGCTCATCTCGAAGGCGCAACAAAGGTGACTTTCGAACTGAAGCTGCTTC 268
 QY 150 TCACAAATTCCTCTTATGATCCAGCTTCTCGAATTCGTTGAAAG-TCTGCCTCCTCC 208
 Db 267 AGGCGAAGACCTCCGAGACTCCGGCTTCTGGACATGCCAAGATCTGCTCCTCC 208
 QY 209 TCCATCTCCCTTCAGGGACCA 229
 Db 207 TCCACCTCTCAGCGGGACCA 187

RESULT 4
 LOCUS F06958 348 bp mRNA EST 20-FEB-1995
 DEFINITION HSC1QCL01 normalized infant brain cDNA Homo sapiens cDNA clone
 C-19C10, mRNA sequence.
 ACCESSION F06958
 VERSION F06958.1 GI:672595
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1. (bases 1 to 348)
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
 Devignes, M.D., Duprat, S., Houllatte, R., Jumeau, M.N., Lamy, B.,
 Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,
 Sebastiani-Kabakthis, C. and Tessier, A.
 IMAGES: molecular integration of the analysis of the human genome
 and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534
 On Sep 21, 1992 this sequence version replaced gi:278976.
 Contact: Genethon
 GenexPress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP50 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr

JOURNAL MEDLINE
 COMMENT

The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavet@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tldb/hgi/ngi.html>)
 Seq primer: M13 Reverse.

FEATURES

source
 1..238
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):180944"
 /db_xref="taxon:9606"
 /clone_lib="HSC172 cells II"
 /cell_type="fibroblast"
 /cell_line="HSC172 (60PDL)"
 /dev_stage="fetal"
 /note="Organ: lung; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI"
 BASE COUNT 48 a 70 c 50 g 66 t 4 others
 ORIGIN

Query Match 15.0%; Score 34.4; DB 32; Length 238;
 Best Local Similarity 63.1%; Pred. No. 3;
 Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 14 TGTCTCTCCATCCAGGAGCGGCGGCTGATGGGTCTGGGCTGCCCTGTGCTC 73
 Db 141 TGTTCCTCCCTCCAGGATCCCTTTGGTGAGTATGTTTCAGGATGACACACAC 200
 QY 74 CTCTTGACCCCTCTGGCAGCTCA 97
 Db 201 CTCTAGATACCTTCAGGACACAC 224

RESULT 7
 AA769782/c
 LOCUS ah71b05.s1 Soares_testis_NHT Homo sapiens cDNA clone 1321041 3',
 DEFINITION mRNA sequence.
 ACCESSION AA769782.1 GI:2821020
 VERSION AA769782
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 412)
 AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 9, 1998 this sequence version replaced gi:936203.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 691 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 402.
 FEATURES
 source
 1..412
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

/clone="1321041"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5',
 TGTACCAATCTGAAGTGGAGGCGGCCCAATTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 111 a 78 c 116 g 107 t
 ORIGIN

Query Match 15.0%; Score 34.4; DB 38; Length 412;
 Best Local Similarity 63.1%; Pred. No. 3.7;
 Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 14 TGTCTCTCCATCCAGGAGCGGCGGCTGATGGGTCTGGGCTGCCCTGTGCTC 73
 Db 328 TGTTCCTCCCTCCAGGATCCCTTTGGTGAGTATGTTTCAGGATGACACACAC 269
 QY 74 CTCTTGACCCCTCTGGCAGCTCA 97
 Db 268 CTCTAGATACCTTCAGGACACAC 245

RESULT 8
 AI632159/c
 LOCUS ts85e12.x1 NCI-CCGAP_GC6 Homo sapiens cDNA clone IMAGE:238094 3',
 DEFINITION mRNA sequence.
 ACCESSION AI632159
 VERSION AI632159.1 GI:4683489
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 458)
 AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 7, 1998 this sequence version replaced gi:3121420.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Glibco
 High quality sequence stop: 439.
 FEATURES
 source
 Location/Qualifiers
 1..458
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:238094"
 /clone_lib="NCI-CCGAP_GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"

128 a	92 c	130 q	108 t

ACCESSION
AI803529

REFERENCE
1 (bases 1 to 470)

Colleague: ROBERT STUR
Tel: (301) 496-1550

FEATURES
Location/Qualifiers
High quality sequence scan: 420.

PAGE COUNT	122 p.	83 c.	121 p.	134 f.
	CONSTRUCTED BY BENTON SQUARE			

analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2000, 13:39:45 ; Search time 479.47 Seconds
(without alignments)
-1513.537 Million cell updates/sec

Title: US-09-092-296-1
Perfect score: 239
Sequence: 1 GCCACCGGACTTCAGTGT.....CCCTTCAGGACCGAGGTCA 239

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl.*

Word size : 0

Number of hits that pass the threshold : 1642386

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_p11.*
- 8: gb_p12.*
- 9: gb_p1.*
- 10: gb_p2.*
- 11: gb_p3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_v1.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_Or.*
- 23: em_Ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_p1.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_v1.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*
- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*

50: gb_p13.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	113.4	47.4	192850	10	AB023048	AB023048 Homo sapi
2	113.4	47.4	200000	10	AP000511	AP000511 Homo sapi
3	113.4	47.4	47323	11	AC005937	AC005937 Homo sapi
4	34.6	14.5	291288	42	AC008878	AC008878 Homo sapi
c 5	34.4	14.4	77322	10	HS0144C9	AL096774 Human DNA
6	34.2	14.3	200000	41	AC006445	AC006445 Homo sapi
7	34	14.2	36676	11	HSJ858B16	AL096768 Human DNA
8	33.8	14.1	62181	44	AC016298	AC016298 Homo sapi
c 9	33.6	14.1	128379	11	HSJ773A18	AL049557 Human DNA
c 10	33.6	14.1	26065	43	AC014817	AC014817 Drosophil
c 11	33.6	14.1	106791	45	AC017080	AC017080 Homo sapi
c 12	33.4	14.0	211190	44	AC016716	AC016716 Homo sapi
c 13	33.4	14.0	191010	45	AC016764	AC016764 Homo sapi
14	33	13.8	1801	7	OS025430	U25430 Oryza sativ
15	33	13.8	97580	11	HS591N18	AL031594 Human DNA
16	32.6	13.6	5280	9	AB011114	AB011114 Homo sapi
c 17	32.4	13.6	179532	42	AC010856	AC010856 Homo sapi
c 18	32.2	13.5	108049	43	AC012294	AC012294 Mus muscu
c 19	31.6	13.2	2803	2	TNU93354	U93354 Thermotoga
c 20	31.6	13.2	103150	8	ATF11111	AL079347 Arabidops
c 21	31.6	13.2	125502	8	ATF1420	AL023094 Arabidops
c 22	31.6	13.2	39503	11	AC005595	AC005595 Homo sapi
c 23	31.6	13.2	727	34	MITM117	L31623 Mitrocoma c
c 24	31.6	13.2	168986	40	AC008013	AC008013 Homo sapi
c 25	31.6	13.2	179622	43	AC009533	AC009533 Homo sapi
c 26	31.6	13.2	208079	44	AC013391	AC013391 Homo sapi
c 27	31.4	13.1	39261	10	AC000387	AC000387 Genomic s
c 28	31.4	13.1	100635	11	AC003104	AC003104 Homo sapi
c 29	31.4	13.1	189146	32	AP000846	AP000846 Homo sapi
c 30	31.4	13.1	203806	33	AL133353	AL133353 Homo sapi
c 31	31.4	13.1	183725	41	AC009558	AC009558 Homo sapi
c 32	31.4	13.1	154791	45	AC015478	AC015478 Homo sapi
c 33	31.2	13.1	177241	11	HS402G11	AL023238 Human DNA
c 34	31.2	13.1	78054	35	AC004296	AC004296 Drosophil
c 35	31.2	13.1	245767	42	AC011607	AC011607 Homo sapi
c 36	31.2	13.1	207043	44	AC011612	AC011612 Homo sapi
c 37	31.2	13.1	203734	44	AC011811	AC011811 Homo sapi
c 38	31	13.0	3509	5	A80086	A80086 Sequence 1
c 39	31	13.0	3510	9	HSX9050	X99050 H.sapiens m
c 40	31	13.0	42244	11	AC004660	AC004660 Homo sapi
c 41	31	13.0	97585	11	HS63G5	Z94160 Homo sapien
c 42	31	13.0	186448	32	AP000785	AP000785 Homo sapi
c 43	31	13.0	172461	33	AC006247	AC006247 Drosophil
c 44	31	13.0	195909	40	AF165124	AF165124 Homo sapi
c 45	31	13.0	120980	41	AC008611	AC008611 Homo sapi

ALIGNMENTS

RESULT 1
AB023048
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:53L9, complete sequence.
ACCESSION AB023048
VERSION AB023048.1 GI:5672603
KEYWORDS HTS.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Bakram,S. and Inoko,H.
TITLE
Molecular Dynamics Of HMC genesis unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
MEDLINE
2007539
REFERENCE
2 (bases 1 to 192650)
AUTHORS
Shina,T. and Takishima,N.
TITLE
Direct Submission
JOURNAL
Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shina, Tokai University School of Medicine, Department of
Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail:tshina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)
FEATURES
Location/Qualifiers
1. 192650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="978SK"
/chromosome="6"
/clone="53L9"
/map="6p21.3"
BASE COUNT 49862 a 44743 c 45833 g 52212 t
ORIGIN
Query Match 47.4%; Score 113.4; DB 10; Length 192650;
Best Local Similarity 95.1%; Pred. No. 2e-25; Mismatches 6; Indels 0; Gaps 0;
Matches 117; Conservative 0;
QY 1 GGCACCGGGAGCTTCAGTGTCTCTCCATCCAGGAGCGGCGGCGGCTCTG 60
Db 177631 GGCACCGGGAGCTTCAGTGTCTCTCCATCCAGGAGCGGCGGCGGCTCTG 177690
QY 61 GGTGCGCCCTTCTCTCTCTTGCACCTCTTGCACCTCACATGGAACGCGCGGTA 120
Db 177691 GGTGCGCCCTTCTCTCTCTTGCACCTCTTGCACCTCACATGGAACGCGGCT 177750
QY 121 TGA 123
Db 177751 AGA 177753
RESULT 2
LOCUS
AP000511 200000 bp DNA PRI 26-SEP-1999
DEFINITION
Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
section 10/20.
ACCESSION
AP000511
VERSION
AP000511.1 GI:5926698
KEYWORDS
Homo sapiens DNA.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS
Shina,S., Tamiya,G., Oka,A. and Inoko,H.
TITLE
Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
JOURNAL
Published Only in DataBase (1999) In press
REFERENCE
2 (bases 1 to 200000)
AUTHORS
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
TITLE
Direct Submission
JOURNAL
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases, Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www.allis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
COMMENT
This sequence is conducted by Tokai University as a JST sequencing

Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
Sequence archive Web site (<http://www.alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www.alis.tokyo.jst.go.jp.

FEATURES
Location/Qualifiers
1. 200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
108774..108864
/standard_name="D6S2088"
/note="SHGC-12985:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735268"
115648..115906
/standard_name="D6S1898"
/note="WI-9418:The location is between each flanking site
of PCR primers."
/db_xref="GDB:678272"
complement(115665..115896)
/note="Cda0vhl0:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443382"
142030..142159
/standard_name="D6S2203"
/note="SHGC-16870:The location is between each flanking
site of PCR primers."
/db_xref="GDB:741215"
complement(142085..142322)
/note="RH18132:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4573021"
complement(142479..142740)
/standard_name="D6S1851"
/note="SHGC-10808:The location is between each flanking
site of PCR primers."
/db_xref="GDB:675281"
complement(164525..164665)
/note="SHGC-3064:The location is between each flanking
site of PCR primers."
/db_xref="GDB:1234116"
complement(169209..169574)
/standard_name="D6S953"
/note="UT5233:The location is between each flanking site
of PCR primers."
/db_xref="GDB:313481"
BASE COUNT 52605 a 47531 c 49366 g 50498 t
ORIGIN
Query Match 47.4%; Score 113.4; DB 10; Length 200000;
Best Local Similarity 95.1%; Pred. No. 2e-25; Mismatches 6; Indels 0; Gaps 0;
Matches 117; Conservative 0;
QY 1 GGCACCGGGAGCTTCAGTGTCTCTCCATCCAGGAGCGGCGGCGGCTCTG 60
Db 109958 GGCACCGGGAGCTTCAGTGTCTCTCCATCCAGGAGCGGCGGCTCTG 110017
QY 61 GGTGCGCCCTTCTCTCTCTTGCACCTCTTGCACCTCACATGGAACGCGGTA 120
Db 110018 GGTGCGCCCTTCTCTCTCTTGCACCTCTTGCACCTCACATGGAACGCGGCT 110077
QY 121 TGA 123
Db 110078 AGA 110080

```

RESULT 3
AC005937
LOCUS Homo sapiens clone UWGC:370M23.002 from 6p21, complete sequence.
AC005937
ACCESSION AC005937
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 47323)
AUTHORS Janer, M., Guillaudeau, T., Vu, Q., Kutayavin, T., Harter, H. and
Geraghty, D.E.
TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
2 (bases 1 to 47323)
AUTHORS Geraghty, D.E. and Olson, M.V.
JOURNAL Direct Submission
Submitted (05-NOV-1998) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhcrc.org)
COMMENT 5': UWGC:370M23.013 (Genbank Accession: AC005530)
3': UWGC:y67cl12 (Genbank Accession: AC004211)
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Double stranded (DS) coverage: 75.5%
DS or two chemistry coverage: 98.9%
Single stranded regions: 3
-----
Sequence Validation:
This sequence has been validated by Multiple Complete Digest
Mapping, comparison of the experimentally derived map digest
fragments with sequence-predicted fragments is given below.
Small fragments below a variable cutoff (approximately 400-600bp)
are not mapped and hence do not appear in the table. There are no
significant remaining discrepancies between the experimental and
predicted values. Uniquely ordered fragment groups are separated
by dashed lines.
-----
BglII HindIII Map Seq Map Seq
1069.11 1050.00 889.55 866.00 30541.40 30653.00
20320.67 20855.00 1050.18 1015.00 3279.08 3231.00
2171.50 2147.00 7268.78 7196.00
2560.20 2531.00 10085.80 9992.00
4335.42 4269.00 11212.78 11131.00
2698.62 2628.00
1927.50 1887.00
-----

```

```

FEATURES
SOURCE
1. 47323
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21"
/sub_clone="UWGC:370M23.002"
/clone_lib="Research Genetics BAC Library"
3647..3932
/repeat_region
/repeat_family="Alu"
complement(4999..5277)
/repeat_region
/repeat_family="Alu"
6285..6372
/repeat_region
/repeat_family="Alu"
complement(6972..7050)
/repeat_region
/repeat_family="MLT1"
7286..7584
/repeat_region
/repeat_family="Alu"
complement(8164..8609)
/repeat_region
complement(21287..21895)
/repeat_region
/repeat_family="Alu"
22715..22957
/repeat_region
/repeat_family="Alu"
25510..25802
/repeat_region
/repeat_family="Alu"
27835..28010
/repeat_region
/repeat_family="MER20"
31295..31594
/repeat_region
/repeat_family="Alu"
33515..33767
/repeat_region
/repeat_family="Alu"
34223..34290
/repeat_region
/repeat_family="MIR"
37372..37648
/repeat_region
/repeat_family="Alu"
38526..38700
/repeat_region
/repeat_family="MER3"
39583..40010
/repeat_region
/repeat_family="Alu"
40046..40156
/repeat_region
/repeat_family="Alu"
43194..43372
/repeat_region
/repeat_family="MER5"
43325
/variation
/variation_note="clonal variation with 3' overlapping clone"
44149
/variation
/variation_note="clonal variation with 3' overlapping clone"
44451
/variation
/variation_note="clonal variation with 3' overlapping clone"
44537
/variation
/variation_note="clonal variation with 3' overlapping clone"
44814
/variation
/variation_note="clonal variation with 3' overlapping clone"
44965
/variation
/variation_note="clonal variation with 3' overlapping clone"
45760
/variation
/variation_note="clonal variation with 3' overlapping clone"
45900
/variation
/variation_note="clonal variation with 3' overlapping clone"
46851
/variation
/variation_note="clonal variation with 3' overlapping clone"
46859
/variation
/variation_note="clonal variation with 3' overlapping clone"
47032
/variation
/variation_note="clonal variation with 3' overlapping clone"
47240..47256
/variation
/variation_note="clonal variation with 3' overlapping clone"

```

/note="clonal variation with 3' overlapping clone -
insertion of 17bp repeat"
BASE COUNT 11556 a 11489 c 12284 g 11994 t
ORIGIN

Query Match 47.4%; Score 113.4; DB 11; Length 47323;
Best Local Similarity 95.18; Pred. No. 1.9e-25;
Matches 117; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCCACCGGACATTCAGTCTCTCCATCCAGGAGCGCAGTGGCCATATGGGCTCG 60
|||||
Db 35117 GCCACCGGACATTCAGTCTCTCCATCCAGGAGCGCAGTGGCCATATGGGCTCG 35176
|||||

QY 61 GCGTCCCTCTGCTCCCTCTTGGACCTCTTGGCAGCTCACATGGAACAGGCGCGGCTA 120
|||||
Db 35177 GCGTCCCTCTGCTCCCTCTTGGACCTCTTGGCAGCTCACATGGAACAGGCTGAGGGCT 35236
|||||

QY 121 TGA 123
||

Db 35237 AGA 35239

RESULT 4

AC008878 291288 bp DNA HTG 31-OCT-1999
LOCUS Homo sapiens chromosome 19 clone C17B-H1_2207023, *** SEQUENCING IN
DEFINITION PROGRESS ***, 160 unordered pieces.

AC008878
VERSION AC008878.2 GI:6165135
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 291288)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished

2 (bases 1 to 291288)
DOE Joint Genome Institute.
TITLE Direct Submission

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 31, 1999 this sequence version replaced gi:5686195.
www.jgi.doe.gov.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 160 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1016: contig of 1016 bp in length
* gap of unknown length
* 1017 1601: contig of 585 bp in length
* gap of unknown length
* 1602 2427: contig of 826 bp in length
* gap of unknown length
* 2428 3083: contig of 656 bp in length
* gap of unknown length
* 3084 3959: contig of 876 bp in length
* gap of unknown length
* 3960 7880: contig of 3921 bp in length
* gap of unknown length
* 7881 8576: contig of 696 bp in length
* gap of unknown length
* 8577 12400: contig of 3824 bp in length
* gap of unknown length
* 12401 13210: contig of 810 bp in length
* gap of unknown length
* 13211 13929: contig of 719 bp in length

* 13930 gap of unknown length
* 16891: contig of 2962 bp in length
* gap of unknown length
* 17522: contig of 631 bp in length
* gap of unknown length
* 18054: contig of 532 bp in length
* gap of unknown length
* 18055 18736: contig of 682 bp in length
* gap of unknown length
* 18737 19619: contig of 883 bp in length
* gap of unknown length
* 19620 20466: contig of 847 bp in length
* gap of unknown length
* 20467 21409: contig of 943 bp in length
* gap of unknown length
* 21410 22300: contig of 891 bp in length
* gap of unknown length
* 22301 23263: contig of 963 bp in length
* gap of unknown length
* 23264 24270: contig of 1007 bp in length
* gap of unknown length
* 24271 25014: contig of 744 bp in length
* gap of unknown length
* 25015 25665: contig of 651 bp in length
* gap of unknown length
* 25666 27090: contig of 1425 bp in length
* gap of unknown length
* 27091 28065: contig of 975 bp in length
* gap of unknown length
* 28066 28837: contig of 772 bp in length
* gap of unknown length
* 28838 32347: contig of 3510 bp in length
* gap of unknown length
* 32348 33118: contig of 771 bp in length
* gap of unknown length
* 33119 33757: contig of 639 bp in length
* gap of unknown length
* 33758 34621: contig of 864 bp in length
* gap of unknown length
* 34622 35157: contig of 536 bp in length
* gap of unknown length
* 35158 36196: contig of 1039 bp in length
* gap of unknown length
* 36197 37333: contig of 1137 bp in length
* gap of unknown length
* 37334 38054: contig of 721 bp in length
* gap of unknown length
* 38055 38762: contig of 708 bp in length
* gap of unknown length
* 38763 39524: contig of 762 bp in length
* gap of unknown length
* 39525 40176: contig of 652 bp in length
* gap of unknown length
* 40177 42851: contig of 2675 bp in length
* gap of unknown length
* 42852 43760: contig of 909 bp in length
* gap of unknown length
* 43761 44291: contig of 531 bp in length
* gap of unknown length
* 44292 44840: contig of 549 bp in length
* gap of unknown length
* 44841 45627: contig of 787 bp in length
* gap of unknown length
* 45628 46298: contig of 671 bp in length
* gap of unknown length
* 46299 47167: contig of 869 bp in length
* gap of unknown length
* 47168 47878: contig of 711 bp in length
* gap of unknown length
* 47879 48711: contig of 833 bp in length
* gap of unknown length
* 48712 48959: contig of 248 bp in length
* gap of unknown length


```

/db_xref="tax

```

006: ПОВРБ -Taty_mn/

```

/chromosome="1"
/map="p34.3-36.11"
/clone="RPI-144C9"
/clone_lib="RFCI-1"
complement(1710..2227)
/note="match: GSS: Em:AQ355665"
complement(1780..2227)
/note="match: GSS: Em:AQ355618"
2858..3151
/note="match: GSS: Em:AQ627537.1"
complement(11439..11580)
/note="match: GSS: Em:AQ331787"
complement(24560..25030)
/note="match: GSS: Em:AQ307331"
complement(24637..25031)
/note="match: GSS: Em:AQ234890"
27999..28360
/note="match: STS: Em:G14651"
29391..29849
/note="match: STS: Em:G06203"
complement(37609..37922)
/note="match: GSS: Em:AQ114724"
47849..48201
/note="match: GSS: Em:AQ553574.1"
63243..63496
/note="match: GSS: Em:AQ142826"
BASE COUNT 21006 a 18901 c 18290 g 19125 t
ORIGIN

```

```

Query Match          14.4%; Score 34.4; DB 10; Length 77322;
Best Local Similarity 63.1%; Pred. No. 1.8;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 18 TGCTCTCCATCCAGGAGCCAGTGCCTATGCGTGGCGTGGCCCTGTCCTC 77
Db 31300 TGTGCTCCCTCCCTCCAGGATCCCTTTGGTGAGTGTGTTCCAGGATGCACCA 31241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 78 CTCTTGACCTCTCTGGCAGCTCA 101
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31240 CTCTAGATACCTTCAGGCAACACA 31217
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 6
AC006445 200000 bp DNA HTG 15-SEP-1999
LOCUS Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 7
DEFINITION unordered pieces.
ACCESSION AC006445
VERSION AC006445.10 GI:5881529
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 200000)
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
Unpublished
2 (bases 1 to 200000)
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
Submitted (28-JAN-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
On Sep 14, 1999 this sequence version replaced gi:5822658.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 1169: contig of 1169 bp in length
* 1170 5668: gap of unknown length
* 5669 12740: contig of 7072 bp in length
* 12741 17239: gap of unknown length
* 17240 27772: contig of 10533 bp in length
* 27773 32270: gap of unknown length
* 32271 67826: contig of 35556 bp in length
* 67827 72325: gap of unknown length
* 72325 108800: contig of 36476 bp in length
* 108801 113298: gap of unknown length
* 113299 160850: contig of 47552 bp in length
* 160851 165348: gap of unknown length
* 165349 200000: contig of 34652 bp in length.
FEATURES             Location/Qualifiers
     source            1..200000
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="4"
BASE COUNT  50025 a 35744 c 35457 g 51783 t 26991 others
ORIGIN

```

Query Match 14.3%; Score 34.2; DB 41; Length 200000;
Best Local Similarity 57.7%; Pred. No. 2.2;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 112 GCGCGGTATGACTTTGCACTGAAGCTGAAGGAGCTTTTCTGACAAATTCCTCTATG 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36776 GACAGGGTATCACTCTGTTGCCAAGCTAAGTCAGTACATGATGACTCACTGTA 36835
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 172 AGTCAGCTCTCTGGAATGCTGAAATTCGCTGCTCTCTCC 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36836 ACCTCAATTCCTGGGATCAAGATGCTCTGCTTGGGCTCC 36879
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
HSJ858B16
LOCUS Homo sapiens DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 858B16 on chromosome 22. Contains the KIA0542 gene and a gene for a novel protein similar to hamster PSSC (Phosphatidylserine Decarboxylase Proenzyme, EC 4.1.1.65). Contains ESTs, GSSs and a putative CpG island, complete sequence.
ACCESSION AL096768
VERSION AL096768.7 GI:5596770
KEYWORDS HTG; CpG island; KIA0542; Phosphatidylserine Decarboxylase Proenzyme; PSSC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 36676)
Barlow,K.
Direct Submission
Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 27, 1999 this sequence version replaced gi:5579004.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

858B16 is from the library RPI-5 constructed at the Roswell Park Cancer Institute by the group of pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.
IMPORTANT: This sequence is not the entire insert of clone 858B16. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 694B4 (AL031255) is at 36572 in this sequence.

FEATURES

source

Location/Qualifiers

```

1..36676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone_lib="RPI-5"
/clone="RP5-858B16"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12717..12948,14414..14546,
14669..14747,15784..15895,15983..16218)
/gene="dJ858B16.1"
/notes="match: ESTs: Em:R53362 Em:Z78408 Em:AA521233
Em:R33904 Em:AI806913 Em:AA158770 Em:AA836285 Em:AA974403
Em:AA594535 Em:AI858196 Em:AI425026 Em:R51102 Em:AA582394
Em:AA829254 Em:AA553345 Em:AI809747 Em:AA846853
Em:AA971234 Em:R54651 Em:AI141022 Em:AA159572 Em:AI361548
Em:AA916775 Em:AI191287 Em:AV031353 Em:AA779744 Em:N64034
Em:AV075750 Em:AV095355 Em:AA865102 Em:AI232370
Em:AA595561 Em:AA088219 Em:K25871 Em:N56896 Em:W90666
Em:AV030271 Em:AV116402 Em:AV122327 Em:AI024249
Em:AA673371 Em:AI130893 Em:AI019759 Em:AI770152
Em:AA149582 Em:N30173 Em:AI342072 Em:R54847 Em:AI126154
Em:AI1816158 Em:AA044074 Em:N54928 Em:AA215582 Em:AA998600
Em:AI623440 Em:AI024124 Em:AA637284 Em:AI824779
Em:AI094871 Em:AI192647 Em:N63934 Em:AV083921"
/evidence="not_experimental"
/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12717..12948,14414..14546,
14669..14747,15784..15895,15983..16218)
/gene="dJ858B16.1"
/notes="match: cDNAs: Em:AB011114"
/evidence="not_experimental"
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
273..16218
/gene="dJ858B16.1"

```

mRNA

gene

CDS

CDS

```

PSTAGPELLLP LSFPMGGAAPARVSAQRATPRDKPPVPSLSASVDPDPLLPGD
FSATRAPGLSTAGSLDELEETQQLHYQTTKNINSCRSQRAOSRLWLELNRE
EPGDEQEVQVQKLEQVEMQQLAELEAQRPQIGACVARIQALQALC"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11528)
/gene="dJ858B16.1"
/notes="original published form: could be artefactual as it
is result of a large intron read-through; match: proteins:
Tr:060289"
/codon_start=1
/evidence="not_experimental"
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
/protein_id="CAB56392.1"
/db_xref="GI:5921489"
/translation="QAIIHAERQLLYRSFWMHQQAARHQEQWQTVACAAHRRGL
KKAFLMRESAQGLRTERTGRVRAAEFMAQLLRWANWSQWRECLARGAEROKLMRAD
LHHQSVLHRLAQAWTYQGRVRAAEFMAQLLRWANWSQWRECLARGAEROKLMRAD
AKKTFQASTHYRTICSKVLQVREAVSYQYRQEDCAIWEAQKVLDRGCLRTWFO
RWDCSRSSAQRLQLERAVQHHRQLLLLEGLARNKTHLQCVKRLLRHQSTQLLAQ
RLSRTCFTQWROQLAARQEQRTVRAALFWAFSLQAKVWATLWLFVLRERKKARLQ
WALQYQGLLOEGATRLRFAASMKASRQQLQAOQVQAAHSLHRAVRRCATLWKOK
VLGRGKPOPLAAIAPSRKVTFFEPPLNRIAGAGDGTLETKRFPQASRPLGALGLAA
EPEHALEI"
620..913
/notes="AluX repeat: matches 1..294 of consensus"
1444..1687
/gene="dJ858B16.1"
/notes="match: GSS: Em:AAQ54234"
3250..3336
/notes="AluJo repeat: matches 2..302 of consensus"
3634..3836
/notes="L2 repeat: matches 2497..2701 of consensus"
4356..4646
/notes="AluX repeat: matches 21..312 of consensus"
4650..4754
/notes="AluSg/x repeat: matches 200..307 of consensus"
4990..5281
/notes="AluX repeat: matches 1..292 of consensus"
5806..6231
/notes="MLTLC repeat: matches 6..460 of consensus"
6529..6668
/notes="FRAM repeat: matches 1..140 of consensus"
6671..6966
/notes="AluJ repeat: matches 1..295 of consensus"
6985..7081
/notes="LTR40a repeat: matches 1..100 of consensus"
7929..8097
/notes="MLR repeat: matches 82..255 of consensus"
9766..10068
/notes="AluJ repeat: matches 1..301 of consensus"
10069..10096
/notes="14 copies 2 mer ta 100% conserved"
10104..10399
/notes="AluJb repeat: matches 1..293 of consensus"
12099..12259
/notes="MLR repeat: matches 46..208 of consensus"
13063..13779
/notes="L2 repeat: matches 1570..2750 of consensus"
15108..15404
/notes="AluSg repeat: matches 1..295 of consensus"
complement(16160)
/gene="dJ858B16.2"
complement(join(16160..17505,18222..18382,18665..18811,
19003..19141,19318..19554,21353..21518,23432..23685,
28473..28494))
/gene="dJ858B16.2"
/notes="match: cDNAs: Em:MG2722 Em:AL050371 Em:AF086277
Em:AF035304 Em:U82163 Em:AB011114; match: ESTs:
Em:AA324939 Em:AA682112 Em:AA166445 Em:AA166440
Em:AA260704 Em:AI019759 Em:AL039237 Em:AA118158
Em:AV083921 Em:AI232370 Em:AU035550 Em:AV122327
Em:AA717291 Em:AA326672 Em:AU051210 Em:AA298268

```

Em:AA144920 Em:AA673371 Em:R00603 Em:AA297077 Em:AA237290
 Em:AA515772 Em:AA297754 Em:AA297076 Em:AA643609
 Em:AI094871 Em:AA298523 Em:AA149582 Em:AA836285
 Em:AV031353 Em:AA297159 Em:AI191287 Em:AA916775 Em:R51102
 Em:N30173 Em:AA521233 Em:AA044074 Em:AI361548 Em:H59648
 Em:AA971234 Em:W00857 Em:AI623440 Em:AA553345 Em:H33993
 Em:AA046738 Em:AA637284 Em:AA151798 Em:W69452 Em:AA297086
 Em:W69536 Em:TI0522 Em:AI024124 Em:H71693 Em:AA302951
 Em:N98280 Em:N63934 Em:AA215768 Em:R94341 Em:AA177542
 Em:R48871 Em:AI342072 Em:W90666 Em:AA255719 Em:AI024249
 Em:AA829254 Em:W69460 Em:AA595561 Em:W69544 Em:T77685
 Em:AA779744 Em:N64034 Em:AV075750 Em:H38339 Em:AV095355
 Em:AV115402 Em:AI425026 Em:AA974403 Em:AA908164
 Em:AA088219 Em:AA298551 Em:N56896 Em:AV030271 Em:AI428571
 Em:AI806913 Em:AI014094 Em:H53175 Em:AI809747 Em:H38743
 Em:H62728 Em:W03262 Em:R52284 Em:AA044485 Em:R82229
 Em:AA369042 Em:AA935183 Em:AI126154 Em:R54847 Em:AA158770
 Em:AA825239 Em:R53399 Em:AI770152 Em:AA594535 Em:AA582394

Query Match 14.2%; Score 34; DB 11; Length 36676;
 Best Local Similarity 54.9%; Pred. No. 2.3;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 QY 19 GTCCTCCATCCAGGAGGAGTGGCCACTATGGGGTCTGGGCTGCCCTGTGCTCC 78
 Db 14428 GGCCTACTCTAGGATAGAGCCCCCGTCCCTCATCCCTGGCCAGTGCCTGACCCCA 14487
 QY 79 TCTTGACCTCTCTGGCAGCTCACATGGAACAGCGCGGGTATGACTTTGCAACTGAAGC 138
 Db 14488 TCTACTCTCTCTGGGAGCTTCTCAGCCACAGCGCTGGCGCTTCAACTSCAGG 14547
 QY 139 TG 140
 Db 14548 TG 14549

RESULT 8
 LOCUS AC016298 62181 bp DNA HTG 24-NOV-1999
 DEFINITION Homo sapiens clone RP11-655M17, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC016298
 VERSION AC016298.1 GI:5467036
 KEYWORDS HTG; HIGS_PHASE0.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 62181)
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Boukhar, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,

Cooke, P., DeArillano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferrel, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehocky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meidlin, J.,

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollars, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wynan, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RW/repeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submission@genome.wi.mit.edu

----- Project Information

Center Project name: I4315

Center Clone name: 655_M_17

 * NOTE: This record contains 78 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 775: contig of 775 bp in length
 * gap of unknown length
 * 776 1583: contig of 808 bp in length
 * gap of unknown length
 * 1584 2371: contig of 788 bp in length
 * gap of unknown length
 * 2372 3185: contig of 814 bp in length
 * gap of unknown length
 * 3186 3980: contig of 795 bp in length
 * gap of unknown length
 * 3981 4778: contig of 798 bp in length
 * gap of unknown length
 * 4779 5584: contig of 806 bp in length
 * gap of unknown length
 * 5585 6392: contig of 808 bp in length
 * gap of unknown length
 * 6393 7189: contig of 787 bp in length
 * gap of unknown length
 * 7190 7994: contig of 805 bp in length
 * gap of unknown length
 * 7995 8802: contig of 808 bp in length
 * gap of unknown length
 * 8803 9603: contig of 801 bp in length
 * gap of unknown length
 * 9604 10400: contig of 797 bp in length
 * gap of unknown length
 * 10401 11206: contig of 806 bp in length
 * gap of unknown length
 * 11207 12024: contig of 818 bp in length
 * gap of unknown length
 * 12025 12840: contig of 816 bp in length
 * gap of unknown length
 * 12841 13637: contig of 797 bp in length
 * gap of unknown length
 * 13638 14440: contig of 803 bp in length
 * gap of unknown length
 * 14441 15234: contig of 794 bp in length
 * gap of unknown length
 * 15235 16031: contig of 797 bp in length
 * gap of unknown length
 * 16032 16836: contig of 805 bp in length
 * gap of unknown length
 * 16837 17615: contig of 779 bp in length
 * gap of unknown length
 * 17616 18421: contig of 806 bp in length
 * gap of unknown length
 * 18422 19233: contig of 812 bp in length
 * gap of unknown length
 * 19234 20042: contig of 809 bp in length
 * gap of unknown length
 * 20043 20856: contig of 814 bp in length
 * gap of unknown length
 * 20857 21657: contig of 801 bp in length
 * gap of unknown length

* 21658 22452: contig of 805 bp in length
* gap of unknown length
* 22463 23262: contig of 800 bp in length
* gap of unknown length
* 23263 24065: contig of 803 bp in length
* gap of unknown length
* 24066 24866: contig of 801 bp in length
* gap of unknown length
* 24867 25673: contig of 807 bp in length
* gap of unknown length
* 25674 26475: contig of 802 bp in length
* gap of unknown length
* 26476 27270: contig of 795 bp in length
* gap of unknown length
* 27271 28075: contig of 805 bp in length
* gap of unknown length
* 28076 28817: contig of 742 bp in length
* gap of unknown length
* 28818 29613: contig of 796 bp in length
* gap of unknown length
* 29614 30416: contig of 803 bp in length
* gap of unknown length
* 30417 31207: contig of 791 bp in length
* gap of unknown length
* 31208 32021: contig of 814 bp in length
* gap of unknown length
* 32022 32825: contig of 804 bp in length
* gap of unknown length
* 32826 33591: contig of 766 bp in length
* gap of unknown length
* 33592 34403: contig of 812 bp in length
* gap of unknown length
* 34404 35213: contig of 810 bp in length
* gap of unknown length
* 35214 36022: contig of 809 bp in length
* gap of unknown length
* 36023 36829: contig of 807 bp in length
* gap of unknown length
* 36830 37623: contig of 794 bp in length
* gap of unknown length
* 37624 38411: contig of 788 bp in length
* gap of unknown length
* 38412 39213: contig of 802 bp in length
* gap of unknown length
* 39214 40010: contig of 797 bp in length
* gap of unknown length
* 40011 40741: contig of 731 bp in length
* gap of unknown length
* 40742 41570: contig of 829 bp in length
* gap of unknown length
* 41571 42298: contig of 728 bp in length
* gap of unknown length
* 42299 42980: contig of 682 bp in length
* gap of unknown length
* 42981 43786: contig of 806 bp in length
* gap of unknown length
* 43787 44584: contig of 798 bp in length
* gap of unknown length
* 44585 45392: contig of 808 bp in length
* gap of unknown length
* 45393 46206: contig of 814 bp in length
* gap of unknown length
* 46207 46991: contig of 785 bp in length
* gap of unknown length
* 46992 47813: contig of 822 bp in length
* gap of unknown length
* 47814 48623: contig of 810 bp in length
* gap of unknown length
* 48624 49437: contig of 814 bp in length
* gap of unknown length
* 49438 50183: contig of 746 bp in length
* gap of unknown length
* 50184 50985: contig of 802 bp in length

* 50986 51800: contig of 815 bp in length
* gap of unknown length
* 51801 52595: contig of 795 bp in length
* gap of unknown length
* 52596 53395: contig of 800 bp in length
* gap of unknown length
* 53396 54205: contig of 810 bp in length
* gap of unknown length
* 54206 54996: contig of 791 bp in length
* gap of unknown length
* 54997 55783: contig of 787 bp in length
* gap of unknown length
* 55784 56587: contig of 804 bp in length
* gap of unknown length
* 56588 57389: contig of 802 bp in length
* gap of unknown length
* 57390 58203: contig of 814 bp in length
* gap of unknown length

Query Match 14.1%; Score 33.8; DB 44; Length 62181;
Best Local Similarity 53.4%; Pred. No 2.7; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 10 GACTTCAGTGTCTCCATCCAGGAGCGCAGTGGCGCACTATGGGTCTGGGCTGCCCC 69
DB 11905 GATTTTAAATTTCTTACCCCTGAAGCTGGGAGTGTCTTCTTCTTGGTGGCTCCCC 11964
OY 70 TTGTCTCTCTTGTACCCCTTGTGAGCTCACAAGGAGGCGGCGGATGACTTTGC 129
DB 11965 TCTTCTTCTTGTGGCCCTGGGAGTGGCAGTCCCTGGAAAAGGTCCTACATCCCTTGC 12024
OY 130 AACTGAGCTGAA 142
DB 12025 GCCTCGGGTTGCA 12037

RESULT 9

HSJ773A18/c

LOCUS HSJ773A18 128379 bp DNA PRI 23-NOV-1999

DEFINITION Human DNA sequence from clone 773A18 on chromosome 1p13.2-21.1
Contains genes for RAS-RELATED PROTEIN RAP-1A (C1KG) (KREV-1
PROTEIN) (GTP-BINDING PROTEIN SMG-P21A) (G-22K), KCND3 (potassium
voltage-gated channel, Shal-related subfamily, member 3), PROBABLE
ATP-DEPENDENT RNA HELICASE P47 HOMOLOG, ESTs, STSs, GSSs and Cpg
Islands, complete sequence.

ACCESSION AL049557

VERSION AL049557.19 GI:5596765

KEYWORDS HTG; C21KG; Cpg Island; G-22K; GTP-BINDING PROTEIN; KCND3; KREV-1

SOURCE PROTEIN; RAP1A; RAS-RELATED PROTEIN RAP1-A; SMG-P21A.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 128379)

AUTHORS Coville G.

TITLE Direct Submission

JOURNAL Submitted (03-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

COMMENT On Jul 27, 1999 this sequence version replaced gi:5579000.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

This sequence is

the entire insert of clone 773A18. This sequence has been finished

according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

773A18 is from the library RPI1-4 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcrPAC2>.

FEATURES

```

source
    1..128379
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="1"
    /map="p13.2-21.1"
    /clone_lib="RPI1-4"
    /clone="RP4-773A18"
    complement(1..988)
    /gene="dJ773A18.3"
    complement(1..988)
    /gene="dJ773A18.3"
    /note="dJ773A18.3 (similar to cytochrome 18); match:
    CDNAS: Em:X12876; match: proteins: Sw:P05784 Sw:P05783"
    /codon_start=1
    /pseudo
    /evidence=not_experimental
    1487..1866
    /note="match: GSS: Em:AQ518605"
    1535..1556
    /note="11 copies 2 mer tt 100% conserved"
    1557..1697
    /note="MIR repeat: matches 76..206 of consensus"
    1872..1977
    /note="U6 repeat: matches 2..107 of consensus"
    1979..2236
    /note="LMC1 repeat: matches 6071..6323 of consensus"
    2237..2239
    /note="AluY repeat: matches 4..299 of consensus"
    2540..2597
    /note="LMC1 repeat: matches 6013..6071 of consensus"
    2915..3217
    /note="AluSx repeat: matches 1..303 of consensus"
    3222..3352
    /note="AluSc repeat: matches 1..131 of consensus"
    3414..3702
    /note="AluSx repeat: matches 13..301 of consensus"
    3855..4028
    /note="MER58C repeat: matches 29..84 of consensus"
    4074..4274
    /note="LMB6 repeat: matches 5411..5607 of consensus"
    4316..4840
    /note="LMB6 repeat: matches 5594..6136 of consensus"
    4841..5150
    /note="AluSp repeat: matches 1..309 of consensus"
    5151..5192
    /note="LMB6 repeat: matches 6136..6175 of consensus"
    5775..5852
    /note="MIR repeat: matches 129..209 of consensus"
    5862..6106
    /note="MIR repeat: matches 5..252 of consensus"
    6123..6223
    /note="MIR repeat: matches 28..138 of consensus"
    6330..6483
    /note="FRAM repeat: matches 1..156 of consensus"
    6665..6727
    /note="31 copies 2 mer aa 77% conserved"
    7100..7429
    /note="AluY5 repeat: matches 1..308 of consensus"
    7838..8216
    /note="L2 repeat: matches 1148..1537 of consensus"
    repeat_region
    8268..8525
    /note="L1MB3 repeat: matches 5916..6183 of consensus"
    repeat_region
    8581..8904
    /note="MER46B repeat: matches 2..236 of consensus"
    repeat_region
    8906..9254
    /note="L1M4 repeat: matches 4791..5146 of consensus"
    repeat_region
    9234..9548
    /note="L1M4 repeat: matches 2658..2976 of consensus"
    repeat_region
    9550..9605
    /note="Trigger3(Golem) repeat: matches 2970..3028 of
    consensus"
    repeat_region
    9606..9702
    /note="L1M4 repeat: matches 4324..4430 of consensus"
    repeat_region
    9730..9809
    /note="AluJ/FRAM repeat: matches 228..305 of consensus"
    repeat_region
    9821..10068
    /note="Trigger3(Golem) repeat: matches 2277..2547 of
    consensus"
    repeat_region
    10057..10171
    /note="Trigger3(Golem) repeat: matches 1..111 of consensus"
    repeat_region
    10172..10563
    /note="L1MEC repeat: matches 2272..2658 of consensus"
    repeat_region
    10564..10876
    /note="MER44A repeat: matches 1..333 of consensus"
    repeat_region
    10877..10971
    /note="L1MEC repeat: matches 2480..2272 of consensus"
    repeat_region
    10972..11047
    /note="L1PA2 repeat: matches 6069..6144 of consensus"
    repeat_region
    11048..11422
    /note="L1MEC repeat: matches 2112..2481 of consensus"
    repeat_region
    11474..11935
    /note="L1MEC repeat: matches 1520..2023 of consensus"
    repeat_region
    11938..12338
    /note="L2 repeat: matches 1958..2404 of consensus"
    repeat_region
    12484..12837
    /note="L2 repeat: matches 2029..2370 of consensus"
    repeat_region
    12960..13258
    /note="AluSg repeat: matches 1..299 of consensus"
    repeat_region
    13555..13859
    /note="AluY repeat: matches 1..305 of consensus"
    repeat_region
    14240..14335
    /note="MIR repeat: matches 38..140 of consensus"
    repeat_region
    14350..14652
    /note="AluYB8 repeat: matches 1..315 of consensus"
    misc_feature
    complement(14862..15301)
    /note="match: ESTs: Em:AI673310"
    repeat_region
    14948..15005
    /note="MIR repeat: matches 171..228 of consensus"
    repeat_region
    15170..15237
    /note="L2 repeat: matches 2682..2749 of consensus"
    misc_feature
    <15266..>15532
    /note="match: ESTs: Em:AA837720"
    repeat_region
    15533..15844
    /note="AluSx repeat: matches 1..312 of consensus"
    misc_feature
    15942..16366
    /note="match: ESTs: Em:AA772504"
    repeat_region
    16244..16342
    /note="MIR repeat: matches 141..262 of consensus"
    misc_feature
    17326..>17651
    /note="match: ESTs: Em:W03890"
    repeat_region
    18127..18623
    /note="L2 repeat: matches 1983..2543 of consensus"
    repeat_region
    18844..19138
    /note="AluSx repeat: matches 1..297 of consensus"
    misc_feature
    complement(19286..>19694)
    /note="match: ESTs: Em:AI281646"
    misc_feature
    19368..19565
    /note="match: ESTs: Em:N76756"
    repeat_region
    19695..19862
    /note="L1MEC repeat: matches 1049..1219 of consensus"
    repeat_region
    19863..20101
    /note="MIR repeat: matches 2..261 of consensus"
    repeat_region
    20102..20186

```

repeat_region /note="L1MEC repeat: matches 1219. .1304 of consensus"
20187. .20493
repeat_region /note="AluJB repeat: matches 1. .311 of consensus"
20494. .20653
repeat_region /note="L1MEC repeat: matches 1304. .1454 of consensus"
20747. .21111
repeat_region /note="MLT1J repeat: matches 111. .516 of consensus"
21166. .21597
repeat_region /note="L1MEC repeat: matches 1509. .1991 of consensus"
21633. .21941
repeat_region /note="AluJB repeat: matches 3. .308 of consensus"
21945. .22194
repeat_region /note="L1M4 repeat: matches 3044. .3318 of consensus"
complement(22225. .22726)
misc_feature /note="match: GSS: Em:AQ180467"
22604. .22705
repeat_region /note="FRAM repeat: matches -2. .120 of consensus"
22908. .23214
repeat_region /note="AluSx repeat: matches 1. .310 of consensus"
23215. .23635

Query Match 14.1%; Score 33.6; DB 11; Length 128379;
Best Local Similarity 48.9%; Pred. NO. 3.3;
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 8 GGGACTTCATGTCCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGCGTCC 67
DB 97429 GTGACCTCTTTTCTCCACCCATGGTCTCAGGAGGAGGAGCGAGTCTTTC 97370

QY 58 CTTGTCCTCTCTTGACCTCTTGCCAGCTCACATGACAGCGCGGGTATGACTTT 127
DB 97369 CTTCTGTAATATGTCCTCCCTGGCAAGCTGCCAACACTGCTCTTGTACAGCTA 97310

QY 128 GCAACTGAAGCTGAAGGAGCTTTCTGACAAATCTCTATGATCCAGCTTCTTGA 187
DB 97309 CTAAGTGAGATGGCCAGTGTATGTATAAATGAGCTTCTGCTGATCCAGGCC 97250

QY 188 ATTG 191
DB 97249 ATGG 97246

RESULT 10
AC014817/c
LOCUS AC014817 26065 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014817
VERSION AC014817.1 GI:6436518
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 26065)
ADAMS, M. and VENTER, J. C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210440 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. .26065
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 7321 a 5960 c 5929 g 6855 t
ORIGIN

Query Match 14.1%; Score 33.6; DB 43; Length 26065;
Best Local Similarity 59.4%; Pred. NO. 3;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 5 ACCGGAGCTTCAGTGTCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGGCT 64
DB 1415 ACAGGCCATGATGTCTTGTCTCCAGGCGCTCATGACCCCTTTTGGCTCCGCT 1356

QY 65 GCCCTTGTCTCTCTTGTACCTCTCTGGCAGCTC 100
DB 1355 GCGCTAAATCCCCACCCCTCTCTTGGCAGCTC 1320

RESULT 11
AC017080/c
LOCUS AC017080 106791 bp DNA HTG 09-DEC-1999
DEFINITION Homo sapiens clone RP11-469G4, *** SEQUENCING IN PROGRESS ***, 34
unordered pieces.
ACCESSION AC017080
VERSION AC017080.1 GI:6554067
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106791)
Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 106791)
Waterston, R.H.
DIRECT SUBMISSION
Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT Center project name: H.NH0469G04.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2077: contig of 2077 bp in length
2078 2097: gap of unknown length
2098 6097: contig of 4000 bp in length
6098 6117: gap of unknown length
6118 8806: contig of 2689 bp in length
8807 8826: gap of unknown length
8827 11555: contig of 2729 bp in length
11556 11575: gap of unknown length
11576 13852: contig of 2277 bp in length
13853 13872: gap of unknown length
13873 16840: contig of 2968 bp in length
16841 16860: gap of unknown length
16861 19354: contig of 2494 bp in length
19355 19373: gap of unknown length
19374 22200: contig of 2827 bp in length
22201 22219: gap of unknown length
22220 26238: gap of unknown length
26239 28539: contig of 2301 bp in length
28540 28558: gap of unknown length
28559 31633: contig of 3075 bp in length
31634 31652: gap of unknown length
31653 35653: contig of 4000 bp in length
35654 35671: gap of unknown length
35672 39671: contig of 4000 bp in length
39672 42315: contig of 2625 bp in length
42316 42334: gap of unknown length

```

* 42335 46334: contig of 4000 bp in length
* 46335 gap of unknown length
* 46354 48690: contig of 2337 bp in length
* 48691 48709: gap of unknown length
* 48710 51147: contig of 2438 bp in length
* 51148 51166: gap of unknown length
* 51167 53808: contig of 2642 bp in length
* 53809 53827: gap of unknown length
* 53828 53891: contig of 2064 bp in length
* 53910 55911: gap of unknown length
* 55911 58044: contig of 2134 bp in length
* 58045 58063: gap of unknown length
* 58064 62063: contig of 4000 bp in length
* 62064 62082: gap of unknown length
* 62083 65082: contig of 4000 bp in length
* 65083 66101: gap of unknown length
* 66102 68113: contig of 2011 bp in length
* 68113 68131: gap of unknown length
* 68132 70857: contig of 2726 bp in length
* 70858 70876: gap of unknown length
* 70877 74876: contig of 4000 bp in length
* 74877 74895: gap of unknown length
* 74896 78895: contig of 4000 bp in length
* 78896 78914: gap of unknown length
* 78915 82915: contig of 4000 bp in length
* 82915 82933: gap of unknown length
* 82934 83811: contig of 2884 bp in length
* 83811 85836: gap of unknown length
* 85837 89837: contig of 4000 bp in length
* 89837 89855: gap of unknown length
* 89856 93499: contig of 3644 bp in length
* 93500 93518: gap of unknown length
* 93519 97519: contig of 4000 bp in length
* 97519 97537: gap of unknown length
* 97538 100201: contig of 2864 bp in length
* 100202 100220: gap of unknown length
* 100221 102772: contig of 2552 bp in length
* 102773 102791: gap of unknown length
* 102792 106791: contig of 4000 bp in length.

FEATURES             Location/Qualifiers
     source            1..106791
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="RP11-469G4"
BASE COUNT    30926 a 22362 c 21610 g 31260 t      633 others
ORIGIN

Query Match          14.1%; Score 33.6; DB 45; Length 106791;
Best Local Similarity 54.5%; Pred. No. 3.3;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 119 TATGACTTTCACCTGAGCTGAGGAGTCTTTCTGCAAAATTCCTCTATGAGTCGAG 178
      || || || || || || || || || || || || || || || || || || || ||
Db 21073 TACCACTGCTCTCTCAAAACCCGTCACATTTCTCTCATTTTCACTATATTAAGCCCAT 21014

QY 179 CTTCTGGAATTCCTGAAAATTCGCTCTCTCCCTCCATCTCCCTTCAGGACCAAGCGTC 238
      || || || || || || || || || || || || || || || || || || || ||
Db 21013 ATTCTTAGAGTGCTCAATATGACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 20954

QY 239 A 239
      |
Db 20953 A 20953

RESULT 12
AC016716/c
LOCUS      AC016716      211190 bp      DNA      HTG      04-DEC-1999
DEFINITION Homo sapiens clone RP11-31213, *** SEQUENCING IN PROGRESS ***, 31
            unordered pieces.
ACCESSION  AC016716
VERSION     AC016716.1 GI:6524349
KEYWORDS   HTG; HTGS_PHASE1.

```

```

SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 211190)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 211190)
Waterston,R.H.
Direct Submission
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
Center project name: H.NH0312103.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2127: contig of 2127 bp in length
* 2128 2145: gap of unknown length
* 2146 4227: contig of 2081 bp in length
* 4227 4244: gap of unknown length
* 4245 6448: contig of 2203 bp in length
* 6448 6465: gap of unknown length
* 6466 9092: contig of 2627 bp in length
* 9093 9110: gap of unknown length
* 9111 11195: contig of 2085 bp in length
* 11196 11213: gap of unknown length
* 11214 14033: contig of 2820 bp in length
* 14034 14051: gap of unknown length
* 14052 17131: contig of 3080 bp in length
* 17132 17149: gap of unknown length
* 17150 20580: contig of 3431 bp in length
* 20581 20598: gap of unknown length
* 20599 22836: contig of 2238 bp in length
* 22837 22854: gap of unknown length
* 22855 25615: contig of 2761 bp in length
* 25616 25633: gap of unknown length
* 25634 27881: contig of 2248 bp in length
* 27882 27899: gap of unknown length
* 27900 32615: contig of 4716 bp in length
* 32616 32633: gap of unknown length
* 32634 36085: contig of 3452 bp in length
* 36086 36103: gap of unknown length
* 36104 40577: contig of 4474 bp in length
* 40578 40595: gap of unknown length
* 40596 44344: contig of 3949 bp in length
* 44345 44362: gap of unknown length
* 44363 48843: contig of 4281 bp in length
* 48844 48861: gap of unknown length
* 48862 54412: contig of 5551 bp in length
* 54413 54430: gap of unknown length
* 54431 59457: contig of 5027 bp in length
* 59458 59475: gap of unknown length
* 59476 66653: contig of 7178 bp in length
* 66654 66671: gap of unknown length
* 66672 72122: contig of 5451 bp in length
* 72123 72140: gap of unknown length
* 72141 79221: contig of 7081 bp in length
* 79222 79239: gap of unknown length
* 79240 88165: contig of 8926 bp in length
* 88166 88183: gap of unknown length
* 88184 97208: contig of 9107 bp in length
* 97209 97308: gap of unknown length
* 97309 106181: contig of 8853 bp in length
* 106182 106199: gap of unknown length
* 106180 116959: contig of 10780 bp in length
* 116960 116977: gap of unknown length

```



```

* 116978 127331: contig of 10354 bp in length
* 127332 127349: gap of unknown length
* 127350 138238: contig of 10889 bp in length
* 138239 138256: gap of unknown length
* 138257 151790: contig of 13534 bp in length
* 151791 151808: gap of unknown length
* 151809 164982: contig of 13174 bp in length
* 164983 165000: gap of unknown length
* 165001 184933: contig of 19895 bp in length
* 184896 184933: gap of unknown length
* 184934 211190: contig of 26277 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..211190
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="RP11-31213"
BASE COUNT  52817 a 52007 c 53208 g 52618 t 540 others
ORIGIN

Query Match      14.0%; Score 33.4; DB 44; Length 211190;
Best Local Similarity 50.0%; Pred. No. 3.9;
Matches 82; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Oy  62  GCTGCCCTTCTCTCTTGGACCTCTTGGAGCTCAGATGGACAGGCGCGGTAT 121
Db 162017 GCAGAACCTTGTCTCTCTCCGACCCACTGGAGCTGGACACAAAGCCCGAGTGGAG 161958

Oy  122  GACTTTGCAACTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTCTATGAGTCCAGCTT 181
Db 161957 CCAGGCTTCTCAGAGGCGATGTGACTCACCTGTGCTCTCCACACCTGCCAGCTG 161898

Oy  182  CCTGGAATTTGTAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 225
Db 161897 CAACCCCTGCAAGAAAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 161854

RESULT 13
AC016764
LOCUS      AC016764 191010 bp DNA HTG 09-DEC-1999
DEFINITION Homo sapiens clone Rpl11-549G13, WORKING DRAFT SEQUENCE, 1 unordered
            pieces
AC016764
AC016764.2 GI:5554041
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 191010)
            Waterston.R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 191010)
            Waterston.R.H.
            Direct Submission
            Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Dec 10, 1999 this sequence version replaced gi:5524275.
            Center project name: H.NH0549G13.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1 191010: contig of 191010 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..191010
                /organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/clone="RP11-549G13"
BASE COUNT  46669 a 48592 c 47516 g 47765 t 468 others
ORIGIN

Query Match      14.0%; Score 33.4; DB 45; Length 191010;
Best Local Similarity 50.0%; Pred. No. 3.9;
Matches 82; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Oy  62  GTCGCCCTTCTCTCTTGGACCTCTTGGAGCTCAGATGGACAGGCGCGGTAT 121
Db 148271 GCAGAACCTTGTCTCTCTCCGACCCACTGGAGCTGGACACAAAGCCCGAGTGGAG 148330

Oy  122  GACTTTGCAACTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTCTATGAGTCCAGCTT 181
Db 148331 CCAGGCTTCTCAGAGGCGATGTGACTCACCTGTGCTCTCTCTCTCTCTCTCTCTCTCT 148390

Oy  182  CCTGGAATTTGCTGAAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 225
Db 148391 CAACCCCTGCAAGAAAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 148434

RESULT 14
OSU25430
LOCUS      OSU25430 1801 bp mRNA PLN 23-MAY-1995
DEFINITION Oryza sativa nucleotide pyrophosphatase precursor, mRNA, complete
            cds.
ACCESSION  U25430
VERSION    U25430.1 GI:818848
KEYWORDS   rice.
SOURCE     Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
            1 (sites)
            Hsing,Y.C., Tsao,C.V., Chow,T., Hsieh,J. and Chen,Z.
            Rice early embryogenesis gene
            Unpublished
            2 (bases 1 to 1801)
            Hsing,Y.C.
            Direct Submission
            Submitted (20-APR-1995) Yue-je C. Hsing, Academia Sinica, Institute
            of Botany, Yien-chu-yuan Rd. Taipei, Taiwan, 11529, Republic of
            China
FEATURES
    source
        Location/Qualifiers
            1..1801
                /organism="Oryza sativa"
                /strain="tainung 67"
                /db_xref="taxon:4530"
                /clone="OSE4"
                /dev_stage="3 to 5 days after anthesis"
                /clone_lib="constructed by Y.C. Hsing et al., Academia
                Sinica, Institute of Botany, Taipei, Taiwan, 11529,
                Republic of China"
                /note="early embryos in seeds"
                40..96
                /note="signal peptide to endoplasmic reticulum"
                40..1479
                /EC_number="3.6.1.9"
                /note="nucleotide phosphodiesterase"
                /codon_start=1
                /product="nucleotide pyrophosphatase precursor"
                /protein_id="AA67067.1"
                /db_xref="GI:818849"
                /translation="MAAAMAAAPPPAAGSDPPPTALLPRTTTTAAAPAPRPSAS
                SRHLTLTAALAVATSYLLILPRLTSLAAPAPAPAAQAQVKLEKPVVILSSGFRF
                GYHKKAATPHRLRIGNGTSAATGLVPIPTLTTPNNHYSTATGLXPSSHGILNNYFPD
                PISGDYFTMSHSPKRWIGLPELWATADQIOAATYTPGSEYKGSWDCPDKYCRHY
                NGVPEERVDAILGYFDLPDSMPQFLYFDLPDQGHQVGPDDPAITAEAVRIDE
                MIGRLIAGLEERGVEFDVNVILVSDHGVGTCDKRLVLDLAPWIKLSEDAWLSMTPE
                LLAIROPDDMSLPDVAKMNEGJSGKVENGEYLMYKEDLPISRLHYADSVRIPIII

```

GLPEGYKVKRSDKNECGAGHGDNAFMSRTIFIAHGPRFEGGRVPSFENVEIY
NTIASIMLEPAPNGSSSFDTILLPSE

mat_peptide 97. .1476
/EC_number="3.6.1.9"
/note="nucleotide phosphodiesterase"
polyA_site 1484. .1488
/product="nucleotide pyrophosphatase"
polyA_site 1572. .1576
/note="putative"
polyA_site 1572. .1576
/note="putative"
BASE COUNT 374 a 527 c 447 g 453 t
ORIGIN

Query Match 13.8%; Score 33; DB 7; Length 1801;
Best Local Similarity 47.5%; Pred. No. 4.1;
Matches 96; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 21 CTCCTCCATCCGAGGAGCGAGTGGCCACTATGGGGTGTGGGTCGCCCTTGTCTCTCTC 80
Db 181 CACCTCTCTCTACCGCGCGCTCGCGTCCGACCTCTACCTCTCTCTCTCTCTCTCTC 240
QY 81 TTGACCTCTCTTGGCAGCTCATGGAACAGCGCGGTATGACTTTGCACTGAAGCTG 140
Db 241 CGACGCGCTCTCTCCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 300
QY 141 AAGGAGTCTTTCTGACAAATCTCTATGAGTCCAGTCTCTGGAATGCTTGAANA 200
Db 301 GAGAGCGGTGTGATCTCATCTCTCGAGCGGTCTCGCTTCCGAGTCCAGCACAAG 360
QY 201 TCTGCT 222
Db 361 GCGCGAGCGCGCGCACATCCACC 382

RESULT 15
LOCUS HS591N18 97580 bp DNA 23-NOV-1999
DEFINITION Human DNA sequence from clone 591N18 on chromosome 22q13.1-13.2.
Contains a COX6B (Cytochrome C Oxidase subunit VIb (EC 1.9.3.1))
pseudogene, ESTs, GSSs and two putative CpG islands, complete
sequence.
ACCESSION AL031594
VERSION AL031594.9 GI:5050980
KEYWORDS HFG; COX6B; CpG Island; Cytochrome C Oxidase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
Steward,C.
Direct Submission
Submitted (25-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquy@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jun 11, 1999 this sequence version replaced gi:4914529.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, SWISSPROT, TrEMBL

This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22

Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
591N18 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: PCYPAC2
IMPORTANT: this sequence is not the entire insert of clone 591N18.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 229A8 (286090) is at 97475 in this
sequence. The true right end of clone 1042K10 (AL022238) is at
34197 in this sequence.

FEATURES
Location/Qualifiers
1..97580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q13.1-13.2"
/clone="RP4-591N18"
/clone_lib="RPC1-4"
1..227
repeat_region
/note="AluY repeat: matches 75. .300 of consensus"
996..1104
repeat_region
/note="FRAM_C repeat: matches 1. .108 of consensus"
1147..1220
repeat_region
/note="AluJ/FRAM repeat: matches 217. .294 of consensus"
1225..1501
repeat_region
/note="AluJo repeat: matches 12. .282 of consensus"
1552..1864
repeat_region
/note="AluY repeat: matches 1. .311 of consensus"
1869..2134
repeat_region
/note="AluJb repeat: matches 21. .275 of consensus"
2942..3234
repeat_region
/note="AluSc repeat: matches 4. .295 of consensus"
4854..5144
repeat_region
/note="AluJb repeat: matches 1. .308 of consensus"
6310..6584
misc_feature
/note="match: GSS AQ038173"
6506..6575
repeat_region
/note="MER3 repeat: matches 134. .209 of consensus"
6585..6665
repeat_region
/note="FRAM_A repeat: matches 27. .116 of consensus"
6673..6694
repeat_region
/note="MER3 repeat: matches 148. .167 of consensus"
6695..6988
repeat_region
/note="AluSg repeat: matches 1. .290 of consensus"
6989..7068
repeat_region
/note="MER3 repeat: matches 58. .148 of consensus"
7497..7761
repeat_region
/note="AluJb repeat: matches 1. .301 of consensus"
7845..8136
repeat_region
/note="AluSx repeat: matches 1. .299 of consensus"
8467..8776
repeat_region
/note="AluY repeat: matches 1. .311 of consensus"
9209..9505
repeat_region
/note="AluSg repeat: matches 1. .295 of consensus"
11780..12007
repeat_region
/note="L1ME1 repeat: matches 5465. .5691 of consensus"
12018..12387
repeat_region
/note="HBXV23 repeat: matches 1184. .1573 of consensus"
12454..12544
repeat_region
/note="L1ME1 repeat: matches 5679. .5763 of consensus"
12545..12843
repeat_region
/note="AluJb repeat: matches 1. .297 of consensus"
12844..13179
repeat_region
/note="L1ME1 repeat: matches 5763. .6159 of consensus"
13240..13373
repeat_region
/note="AluJo repeat: matches 1. .136 of consensus"
13408..13612
repeat_region
/note="AluSg/x repeat: matches 96. .293 of consensus"
13625..13924
repeat_region
/note="AluSx repeat: matches 1. .301 of consensus"
13936..14230
repeat_region

```
repeat_region /note="AluSg repeat: matches 1. .297 of consensus"
14237. .14456
repeat_region /note="MIR repeat: matches 11. .261 of consensus"
14729. .14785
repeat_region /note="MIR repeat: matches 202. .261 of consensus"
15147. .15467
repeat_region /note="L2 repeat: matches 2370. .2690 of consensus"
15625. .15910
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
16000. .16300
repeat_region /note="AluSg repeat: matches 1. .291 of consensus"
16422. .16732
repeat_region /note="AluJb repeat: matches 1. .311 of consensus"
16733. .16813
repeat_region /note="L2 repeat: matches 2667. .2746 of consensus"
complement(16814. .17293)
misc_feature /note="match: GSSs AQ377982 AQ380087"
16907. .17100
repeat_region /note="MIR repeat: matches 2. .197 of consensus"
18063. .18376
repeat_region /note="AluJb repeat: matches 1. .307 of consensus"
18462. .18537
repeat_region /note="38 copies 2 mer tt 78% conserved"
18540. .19522
misc_feature /note="SVA repeat: matches 2. .954 of consensus"
19061. .20976
repeat_region /note="random repeat. Some base discrepancies edited.
Assembly is consistent with Restriction Digest."
19115. .20959
misc_feature /note="CpG island"
/evidence=not_experimental
19540. .20001
repeat_region /note="SVA repeat: matches 521. .954 of consensus"
20019. .20360
repeat_region /note="SVA repeat: matches 521. .845 of consensus"
20361. .20854
misc_feature /note="SVA repeat: matches 519. .994 of consensus"
complement(20855. .20945)
repeat_region /note="match: GSS AQ473058"
21493. .21658
repeat_region /note="AluSx repeat: matches 1. .165 of consensus"
21793. .22086
repeat_region /note="AluSx repeat: matches 1. .297 of consensus"
22379. .22534
repeat_region /note="FRAM repeat: matches -1. .162 of consensus"
23688. .24049
repeat_region /note="MLT1A1 repeat: matches 1. .365 of consensus"
24100. .24408
repeat_region /note="AluYb8 repeat: matches 1. .305 of consensus"
25227. .25536
repeat_region /note="AluJb repeat: matches 1. .310 of consensus"
25556. .25779
repeat_region /note="AluJb repeat: matches 58. .297 of consensus"
25834. .26147
repeat_region /note="AluSg repeat: matches 1. .307 of consensus"
26148. .26192
repeat_region /note="MIR repeat: matches 187. .230 of consensus"
26377. .26688
repeat_region /note="AluSx repeat: matches 2. .311 of consensus"
26884. .27010
repeat_region /note="FLAM_C repeat: matches 1. .133 of consensus"
30237. .30344
repeat_region /note="MIR repeat: matches 103. .206 of consensus"
31467. .31777
repeat_region /note="AluSg repeat: matches 1. .310 of consensus"
31963. .32080
repeat_region /note="AluJb repeat: matches 1. .118 of consensus"
32118. .32408
repeat_region /note="AluJb repeat: matches 1. .292 of consensus"
32623. .32927
repeat_region /note="AluJb repeat: matches 1. .300 of consensus"
33051. .33354
repeat_region /note="AluSg repeat: matches 1. .304 of consensus"
```

```
repeat_region 33381. .33519
/note="L2 repeat: matches 2572. .2709 of consensus"
repeat_region 33560. .33777
/note="AluSx repeat: matches 1. .219 of consensus"
repeat_region 34284. .34597
/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 35062. .35370
/note="AluJb repeat: matches 1. .309 of consensus"
repeat_region 35600. .35753
/note="MIR repeat: matches 2. .230 of consensus"
repeat_region 35782. .35854
/note="MER21B repeat: matches 720. .790 of consensus"
repeat_region 35912. .36626
/note="MER21B repeat: matches 1. .713 of consensus"
repeat_region 36660. .36969
/note="AluY repeat: matches 1. .308 of consensus"
repeat_region 37152. .37273
/note="AluSg/x repeat: matches 185. .302 of consensus"

Query Match 13.8%; Score 33; DB 11; Length 97580;
Best Local Similarity 55.3%; Pred.No. 5;
Matches 63; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 91 TTGCGACCTCAGTGAACAGCGCCGGGTATGACTTTCGAACCTGAAGCTGAAGGAGTCTT 150
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2606 TTCCCACTCCACATCCCCAGGTTTAGTAAGACTTTTCAAAAGCAAGGTAGGTGTAAC 2665

QY 151 TTCGTACAATTCCTCCTATAGTCCAGCTTCTCGAATTCCTTGAAATTCG 204
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2666 TGTAAGAGCTTCATGTTATAGTACCGCAACTTTGTATTTCCTGAAGTCTGTG 2719
```

Search completed: March 21, 2000, 14:39:02
Job time: 3557 sec

Result No.	Query	Score	Match	Length	DB	ID	Description
1	227	95.0	439	1	V84366	Human stomach carc	
2	34.4	14.4	2923	1	X04326	Human secreted pro	
3	32	13.4	1128	1	V28846	Mouse coxsackievir	
4	31	13.0	3510	1	T18696	RAP-1 radiation pr	
c	5	30.4	12.7	580	1	T02964	Cell-cycle regulat
c	6	30.4	12.7	580	1	T02964	Mouse coxsackievir
7	29.8	12.5	1301	1	V50430	EST clone AR34, Ne	
8	29	12.1	2259	1	O05106	Sequence encoding	
9	29	12.1	1110	1	N60409	Human apolipoprote	
10	29	12.1	1110	1	N50450	Sequence encoding	
c	11	29	12.1	2095	1	O57953	Human betal, 6'-N-ac
c	12	29	12.1	597	1	O60939	ApoB4L protease cd
c	13	29	12.1	936	1	O63101	ApoE4Lx2 protease
c	14	29	12.1	936	1	T18068	Human ApoE4 cDNA,
c	15	29	12.1	936	1	T18070	Human ApoE4Lx2 cdn
16	29	12.1	1157	1	T06957	Human apolipoprote	
c	17	29	12.1	2432	1	T77045	cdNA encoding glyco
c	18	29	12.1	110000	1	V30458_1	Continuation (2 of
c	19	29	12.1	110000	1	V30459_1	Continuation (2 of
c	20	28.8	12.1	1785	1	T66986	Human brain-specifi
21	28.6	12.0	683	1	O63794	Bovine trypsin gen	
22	28.6	12.0	701	1	O63795	Bovine trypsinogen	
23	28.4	11.9	5359	1	O50996	Rat NMDA receptor	
24	28	11.7	6030	1	V19007	Rattus norvegicus	
c	25	27.6	11.5	3088	1	T05628	ADP ribosylation f
26	27.6	11.5	1294	1	T60587	Protein cognate of	
c	27	27.6	11.5	5102	1	V10261	Wnt-10b gene. Expr
28	27.2	11.4	2122	1	T49318	Human COL4A6 gene.	
c	29	27.2	11.4	5535	1	X21355	Human BAI1 gene. N
30	27.2	11.4	2161	1	X03034	Human IL-1ra BAC c	
31	27.2	11.4	1901	1	X02998	Human IL-1ra BAC c	
32	27	11.3	384	1	O59126	Human brain expres	
c	33	27	11.3	1833	1	X46339	Serine threonine k
c	34	27	11.3	1593	1	X06832	Disease associated
35	26.8	11.2	2554	1	O71243	Serine protease fo	
36	26.8	11.2	2703	1	O83695	New TGF-beta faml	
37	26.8	11.2	2341	1	T31601	Cartilage-derived	
38	26.8	11.2	2341	1	T31601	Human bone morpho	
39	26.8	11.2	2703	1	T59405	DNA encoding human	

QY	1	GGCCACGGGACTTCAGTGTCTCTCATCCACGAGCGCAGTGGCCACTATGGGCTGTG	60
Db	25	GGCCACGGGACTTCAGTGTCTCTCCATCCACGAGCGCAGTGGCCACTATGGGCTGTG	84
QY	61	GGCTGCCCTTGTCTCTCTGTGACCCCTCCTTGGCAGCTCACATGGAACAGGGCGCGGTA	120
Db	85	GGCTGCCCTTGTCTCTCTGTGACCCCTCCTTGGCAGCTCACATGGAACAGGGCGCGGTA	144
QY	121	TGACTTTTGCACTGAACTGAAAGTCTGAAGGAGTCTTTTTCACAAATTCCTCCTATGAGTCCAGCT	180
Db	145	TGACTTTTGCACTGAACTGAAAGTCTTTTTCACAAATTCCTCCTATGAGTCCAGCT	204
QY	181	TCCTGGAAATGGTGTGAAAA-NTCGTGCTCCTCCTTCATCTCCCTTCAGGAGCACAGCTCA	239

```

Db 205 TCTCGAATGCTGAAGCTCTGCTCTCTCCATCTCCATCTCCAGGACCGGTCA 264
|||||
CC which tissues they are most highly expressed in (see X04311 for described
CC uses).
SQ Sequence 2923 BP; 637 A; 818 C; 753 G; 707 T;

Query Match 14.4%; Score 34.4; DB 1; Length 2923;
Best Local Similarity 63.1%; Pred. No. 0.13; 31; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 18 TGTCTCTCCATCCAGGAGCGGACGCTATGGGCTCTGGGCTCGCCCTGTGCTC 77
||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 955 TGTGCTCTCCCTCCAGGACGCTCTTTGGTGAGTATGCTGTTTCAGGATGCACCCACCAC 1014
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 78 CTCCTGACCTCTCTGGCAGCTCA 101
||||| ||| ||||| |||||
Db 1015 CTCAGATACCTTCAGGCAACACA 1038
||||| ||| ||||| |||||

RESULT 3
V28846
ID V28846 standard; DNA: 1128 BP.
AC V28846;
DT 03-AUG-1998 (first entry)
DE Mouse coxsackievirus and adenovirus receptor encoding DNA.
KW Mouse; coxsackievirus; adenovirus; receptor; CAR; cardiac infection;
KW myocarditis; pericarditis; dilated cardiomyopathy; meningoencephalitis;
KW pancreatic infection; acute pancreatitis; gastrointestinal tract;
KW diabetes mellitus; ss.
OS Mus sp.
PH Key
FT Location/Qualifiers
FT 1..1128
FT /*tag= a
FT /product= "coxsackievirus and adenovirus receptor"
FT /transl_except= (pos:1096..1098,aa:Xaa)
FT /note= "Xaa= a stop codon, the sequence is shown as
FT continuing but gets unclear"
FT
PN W09811221-A2.
PD 19-MAR-1998.
PF 12-SEP-1997; U16189.
PR 13-SEP-1996; US-026100.
PA (DAND ) DANA FARBER CANCER INST INC.
PI Bergelson JM, Finberg RW, Horowitz MS;
DR WPI; 98-207384/18.
PT DNA encoding coxsackie virus and adenovirus receptor - useful for
PT modulation of e.g. cardiac, pancreatic or gastrointestinal infection
PS Disclosure; Fig 14; 104pp; English.
CC The present sequence encodes mouse coxsackievirus and adenovirus
CC receptor (CAR). The present invention also describes: (1) a method for
CC which modulates CAR expression comprising contacting the cell with an agent
CC modulating CAR expression comprising activity or CAR nucleic acid expression,
CC such that a cell associated activity is altered relative to a cell,
CC associated activity of the cell in the absence of the agent; and (2) a
CC method for detecting the presence of CAR in a biological sample
CC comprising contacting a biological sample with an agent capable of
CC detecting CAR protein or mRNA such that the presence of CAR is detected.
CC Modulation of CAR is useful for treatment of cardiac infection, e.g.
CC myocarditis, pericarditis or dilated cardiomyopathy, or infection of the
CC central nervous system, e.g. a non-specific febrile illness or
CC meningoencephalitis, or pancreatic infection, e.g. acute pancreatitis,
CC or infection of the respiratory or gastrointestinal tract or childhood
CC onset diabetes mellitus. Probes derived from CAR nucleic acids are
CC useful for hybridisation assays, and antibodies raised against CAR
CC protein are useful for blocking CAR expression. Cell-free assays which
CC include combining CAR protein and a candidate/test compound are useful
CC in screening for drugs which interact with CAR protein.
CC
SQ Sequence 1128 BP; 297 A; 286 C; 288 G; 257 T;

Query Match 13.4%; Score 32; DB 1; Length 1128;
Best Local Similarity 62.5%; Pred. No. 0.53;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

DE Human secreted protein gene 16 clone HSHBQ68.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens
KW W09856804-A1.
PN W09856804-A1.
PD 17-DEC-1998.
FF 11-JUN-1998; U12125.
PR 02-OCT-1997; US-061060.
PR 13-JUN-1997; US-049547.
PR 13-JUN-1997; US-049548.
PR 13-JUN-1997; US-049549.
PR 13-JUN-1997; US-049550.
PR 13-JUN-1997; US-049606.
PR 13-JUN-1997; US-049607.
PR 13-JUN-1997; US-049608.
PR 13-JUN-1997; US-049609.
PR 13-JUN-1997; US-049610.
PR 13-JUN-1997; US-049611.
PR 13-JUN-1997; US-050566.
PR 13-JUN-1997; US-050901.
PR 13-JUN-1997; US-052989.
PR 08-JUL-1997; US-051919.
PR 18-AUG-1997; US-055984.
PR 12-SEP-1997; US-058665.
PR 12-SEP-1997; US-058666.
PR 12-SEP-1997; US-058669.
PR 12-SEP-1997; US-058750.
PR 12-SEP-1997; US-058971.
PR 12-SEP-1997; US-058972.
PR 02-OCT-1997; US-058975.
PR 02-OCT-1997; US-060834.
PR 02-OCT-1997; US-060841.
PR 02-OCT-1997; US-060844.
PR 02-OCT-1997; US-060865.
PR 02-OCT-1997; US-061059.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Lafleur DW,
PI Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P,
PI Yu G;
DR WPI; 99-080881/07.
DR P-PDB; W78141.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 180-182; 380pp; English.
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X04302) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 86 novel genes and their fragments (nucleic acid
CC sequences: X04311-X04310; amino acid sequences W78126-W78225) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 86 polynucleotides, based on
```

```

QY 33 AGGAGCGCAGTGGCCACTATGAGGGTCTGGGCTGCGCCCTTCTCTCTCTGACCTCTCTT 92
DB 712 ACATCGCGGGCGCGCGCTATAGGAGCGTGTGGCCCTTGTCTCATCGGGGCCATCTCT 771
QY 93 GGCAGCTCACATGGACAGG 112
DB 772 TTCTGCTGTACAGGAAGG 791

RESULT 4
TI8696
ID TI8696 standard; cDNA; 3510 BP.
AC TI8696;
DE 05-JUL-1996 (first entry)
DE RAP-1 radiation protecting checkpoint protein cDNA.
KW RAP-1; radiation protecting checkpoint protein; apoptosis;
KW cell death; cancer; diagnosis; therapy; radiotherapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 176..2122
FT FT /*tag= a
FT FT misc_rna 31077..3510
FT FT /*tag= b
FT FT misc_difference 3648
FT FT /*label= cDNA3
FT FT /*tag= b
FT FT /*note= "base n at position 3648 is not identified
FT FT in the specification"
PN WO9611562-A2.
PD 25-APR-1996.
PF 11-OCT-1995; U12445.
PR 11-OCT-1994; IL-111238.
PA (SHOS/) SHOSHAN H Z.
PA (UVR-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI Ganaani D.
DR WPI: 96-221643/22.
DR P-PSDB: R94906.
PT New gene encoding a radiation protecting checkpoint protein - useful
PT for diagnosis and treatment of cancer and other diseases involving
PT abnormal apoptosis
PS Claim 2; Fig 3; 29pp; English.
CC A cDNA clone (TI8696) codes for a novel radiation-protecting human
CC checkpoint protein (R94906), designated RAP-1, involved in
CC regulation of cell cycle progression and/or apoptosis. It was
CC isolated from a human cDNA library established in an immortalised
CC xeroderma pigmentosa cell line, GM2096 (XPIM1). A cDNA (cDNA3) contg.
CC a 462 bp insert complemented UV sensitivity and was used to screen a
CC lambda-gt10-K562 human cDNA library. The insert represented part of
CC the 3510 bp RAP-1 mRNA. The human checkpoint gene can be used as a
CC marker for early detection and prevention of tumour progression or
CC can be used to design new drugs that enhance radiotherapy or
CC chemotherapy of diseases involving abnormal apoptosis.
SQ Sequence 3510 BP; 974 A; 795 C; 848 G; 892 T;

Query Match 13.0%; Score 31; DB 1; Length 3510;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

OY 1 GGCACCGGGACTTCAGTGTCTCTCCATCCAGGAGCGCGCTGCGCCACTATGGGTCTG 60
DB 1757 GCCCAGCTGTGACCCGCTCCCTCCATGGGAGACGAGAAAGATACATCTCTA 1816
QY 61 GGTGCCCCCTGTCTCTCTGTGACCTCTCTTGCGAGCTCACATGGAACAGGCGGGTA 120
DB 1817 TCCTCTCTCTGGTATACCTCTCTGTGACTCTCTCAAGAAACAAAGAAAAGGAGGAT 1876
QY 121 TGACTTTGCACTGAAGCTGAAG 143
DB 1877 CTAGTGGCAGCTTAACGAGG 1899

RESULT 5
PT DNA encoding human receptor for adenovirus C and coxsackievirus B -
PT for preventing and treating viral infection and rendering cells

```

```

T02964/c
ID T02964 standard; cDNA; 580 BP.
AC T02964;
DE 01-MAR-1996 (first entry)
DE Cell-cycle regulatory protein p15 cDNA.
KW Cell-cycle regulatory protein p15; cyclin-dependent kinase inhibitor;
KW CCR; gene therapy; transgenic animal; cancer; cell proliferation;
KW ss; ds.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 91..483
FT FT /*tag= a
FT FT WO9528483-A1.
FT PD 25-OCT-1995.
FT PF 14-APR-1995; U04636.
FT PR 14-APR-1994; US-227371.
FT PR 25-MAY-1994; US-248812.
FT PR 14-SEP-1994; US-306511.
FT PR 29-NOV-1994; US-346147.
FT PA (COLD-) COLD SPRING HARBOR LAB.
FT PI Beach DH, Demetrick DJ, Hannon GJ, Serrano M;
FT WPI: 95-373798/48.
FT P-PSDB: R85118.
FT New cell cycle regulating proteins bind to cyclin dependent kinase -
FT and related nucleic acids, antibodies etc., used in diagnosis and
FT therapy of abnormal cell proliferation, degeneration etc.
FT PS Claim 43; Page 81-82; 109pp; English.
CC cDNA (T02964) coding for the mouse cell-cycle regulatory (CCR)
CC protein p15 (R85118) was isolated from an embryonal carcinoma
CC library using a probe based on a mouse CCR pl3.5 cDNA (T02965).
CC The isolated cDNA can be used; to detect mutations in CCR
CC genes that lead to cell proliferation; to breed transgenic
CC animals to study cellular disorders involving CCR allele
CC mutation/misexpression; and to correct CCR-deficient cells
CC (gene therapy).
SQ Sequence 580 BP; 109 A; 187 C; 197 G; 87 T;

Query Match 12.7%; Score 30.4; DB 1; Length 580;
Best Local Similarity 57.3%; Pred. No. 1.3;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 39 GCAGTGGCCACTATGAGGGTGTGGGTGCGCCCTTGTCTCTCTTGTGACCTCTCTGCGCAGC 98
DB 128 GCGGTGGCCAGGCGCGCTACTGTCTGCGCCCAACATGCGCTTGTCCCGGTCTGTGCG 69
QY 99 TCACATGGCAACAGGCGCGGTATGACTTTTCAACTG 134
DB 68 AGAATGTTCTTCGCGCGCGGTGAGATTCTCTACAG 33

RESULT 6
V50430
ID V50430 standard; cDNA; 1301 BP.
AC V50430;
DE 07-DEC-1998 (first entry)
DE Mouse coxsackievirus and Ad2 and Ad5 receptor (MCAR) cDNA.
KW MCAR; coxsackievirus receptor; CVB; adenovirus; Ad2 receptor;
KW Ad5 receptor; mouse; infection; vaccine; gene therapy; ds.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 157..1215
FT FT /*tag= a
FT FT WO9833819-A1.
FT PD 06-AUG-1998.
FT PF 30-JAN-1998; U01724.
FT PR 30-JAN-1997; US-036986.
FT PA (UNIV ) UNIV NEW YORK STATE.
FT PI Philipson L, Tomko RP;
FT WPI: 98-437397/37.
FT P-PSDB: W69698.
PT DNA encoding human receptor for adenovirus C and coxsackievirus B -
PT for preventing and treating viral infection and rendering cells

```

PT susceptible to transformation by adenoviral vectors in gene therapy
 PS Claim 2; Page 68-70; 88pp; English.
 CC This cDNA molecule codes for mouse MCAR protein (see W69698) that
 CC serves as a cellular receptor for adenoviruses of the serotypes 2
 CC and 5 (subgroup C) and for the group B coxsackieviruses (CVB). It
 CC was obtained by screening a lambda phage expression cDNA library
 CC with antiserum containing antibodies specific for mouse CVB
 CC binding protein p46, and was used to identify human MCAR cDNA
 CC (see V50429). The invention also provides host cells transformed
 CC with DNA molecules encoding HCAR or MCAR and methods of producing
 CC the recombinant proteins or their derivatives. These proteins,
 CC their extracellular domains, as well as oligopeptides (see
 CC W69699-708) which bind virus, are also provided. Isolated HCAR or
 CC MCAR proteins or their fragments or variants are used to prevent
 CC or treat virus infections. Expressing the DNA in cells which lack
 CC these viral receptors renders the cells susceptible to
 CC transformation by adenoviral vectors carrying genes for gene
 CC therapy.
 SQ Sequence 1301 BP; 355 A; 327 C; 322 G; 297 T;
 Query Match 12.7%; Score 30.4; DB 1; Length 1301;
 Best Local Similarity 61.2%; Pred. No. 1.8;
 Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 33 AGGAGCGCAGTGGCCACTATGGGTGGGCTGGCCCTTCTCTCTGACCTCTCTT 92
 DB 868 ACCATCGCGGGCGGCTCATAGGAGCGCTGCTGGCCCTTCTCTCTGACCTCTCT 927
 QY 93 GGCAGCTCACAGGACAGG 112
 DB 928 TTCGCTGTCACAGGAAACG 947
 RESULT 7
 ID V86365/C
 AC V86365;
 DE EST clone AR34
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS Homo sapiens.
 PN WO9845435-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06954.
 PR 10-APR-1997; US-835913
 PA (GENY) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI; 99-070076/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 210; 633pp; English.
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoietic regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC therapy. The EST sequences are also stated to be useful for gene
 SQ Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;

Query Match 12.5%; Score 29.8; DB 1; Length 291;
 Best Local Similarity 70.2%; Pred. No. 1.6;
 Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 119 TATGACTTCCAACTGAAGGAGTCTTTCTGACAAATTCCTCCATGAGTC 175
 DB 257 TTGATTTTGAACAGAGCTGAGTATTTTCTTACATTTTCTTAAATATC 201
 RESULT 8
 Q05106
 ID Q05106 standard; DNA; 2259 BP.
 AC Q05106;
 DT 02-NOV-1990 (first entry)
 DE Sequence encoding tumour-associated antigen GA733-1.
 KW Pancreatic carcinoma; GA733-1; cancer; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 307..1275
 FT /*tag= a
 FT PF-376746-A.
 PN 4-JUL-1990.
 PD 23-DEC-1989; 313687
 PR 29-DEC-1988; US-291593.
 PA (WIST-) Wistar Inst.
 PI Linnenbach A;
 DR WPI; 90-203091/27.
 DR P-PSDB; R05711.
 PT Tumour-associated antigen, GA733-2 -
 PT Expressed in pancreatic carcinoma cells, used for tumor
 PT immuno-therapy.
 PS Claim 2; Fig 2A; 12pp; English.
 CC The peptide is produced by tumor cells, and Abs raised to the
 CC peptide may be used in diagnosis and therapy of human tumours.
 SQ Sequence 2259 BP; 460 A; 648 C; 639 G; 512 T;
 Query Match 12.1%; Score 29; DB 1; Length 2259;
 Best Local Similarity 52.0%; Pred. No. 6;
 Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 21 CTCCTCCATCCAGGAGCCGACATGCGCTATGGGTCTGGCTGCCCTGTCTCTC 80
 DB 1120 CGCCTACCGCGCGCTCATCGCCCTCATCGTGTGTGTGTGTGTGTGTGTGT 1179
 QY 81 TTGACCCCTCTTGGCAGCTCACATGGAACAGCGCGGTATGACTTTGCAACTGAAGCTG 140
 DB 1180 ATGGCGGCTCTGTGTATCCACCAACCGAGAAAGTCGGGAAGTACAGAAGGTGGAGATC 1239
 QY 141 AAGGA 145
 DB 1240 AAGGA 1244
 RESULT 9
 N60409
 ID N60409 standard; DNA; 1110 BP.
 AC N60409;
 DT 01-JAN-1980 (first entry)
 DE Human apolipoprotein-E.
 KW Apolipoprotein-E; Hyperlipidemia; arteriosclerosis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 15..968
 FT /*tag= a
 FT PN A08547513-A.
 PD 24-APR-1986.
 PF 17-SEP-1985; 047513
 PR 11-JUN-1985; JP-126989.
 PA (MITU) MITSUBISHI CHEM IND KK.
 DR WPI; 86-150217/24.
 DR P-PSDB; P60507.

PT New DNA sequence coding for human apolipoprotein-E - and
 PT expression vectors and transformed cells contg. it
 PS Disclosure; Fig 2; 45pp; English.
 CC The encoded protein is used to treat subjects who are deficient in
 CC apolipoprotein-E (or who produce abnormal forms of this molecule)
 CC and therefore are likely to suffer from hyperlipidemia, resulting in
 CC arteriosclerosis. It can also be used to raise antisera for
 CC detecting the protein deficiency or production of abnormal forms.
 SQ Sequence 1110 BP; 198 A; 354 C; 415 G; 143 T;

Query Match 12.1%; Score 29; DB 1; Length 1110;
 Best Local Similarity 54.1%; Pred. No. 4.7;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 10 GACTTCAGTCTCTCTCCATCCAGGAGCGCAGTGGCAGTATGGGTGGGTGCCCC 69
 Db GAGCGCGGCTCAGCGCCATCCGAGCGCTGGGGCCCTGTGTGAACAGGCGCGGTG 638

QY 70 TTGTCTCTCTTGACCTCTCTGGCAGCTCACATGAACAGGCGCGG 118
 Db CGGCGCGCAGTGTGGTCTCTGGCGGCGCAGCGCTACAGGAGCGG 687

RESULT 10
 N50450
 ID N50450 standard; DNA; 1110 BP.
 AC N50450;
 DT 09-JAN-1992 (first entry)
 DE Sequence encoding human apolipoprotein E.
 KW Hyperlipaemia; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 15..988
 FT mat_peptide 69..965
 FT /*tag= a
 FT /*tag= b
 FT J60118189-A.
 PN 25-JUN-1985.
 PF 29-NOV-1983; 224980.
 PR 29-NOV-1983; JP-224980.
 PA (MITU) MITSUBISHI CHEM IND KK.
 DR WPI; 85-188003/31.
 DR P-PSDB; P51204.
 PT DNA fragment - contg. DNA which codes human apolipoprotein E for
 PT treatment of hyperlipaemia.
 PS Claim 3; Page 484; 8pp; Japanese.
 CC The sequence may be used to produce the apolipoprotein E, useful in the
 CC treatment of hyperlipaemia.
 SQ Sequence 1110 BP; 198 A; 355 C; 414 G; 143 T;

Query Match 12.1%; Score 29; DB 1; Length 1110;
 Best Local Similarity 54.1%; Pred. No. 4.7;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 10 GACTTCAGTCTCTCTCCATCCAGGAGCGCAGTGGCAGTATGGGTGGGTGCCCC 69
 Db GAGCGCGGCTCAGCGCCATCCGAGCGCTGGGGCCCTGTGTGAACAGGCGCGGTG 638

QY 70 TTGTCTCTCTTGACCTCTCTGGCAGCTCACATGAACAGGCGCGG 118
 Db CGGCGCGCAGTGTGGTCTCTGGCGGCGCAGCGCTACAGGAGCGG 687

RESULT 11
 Q57953/C
 ID Q57953 standard; cDNA; 2095 BP.
 AC Q57953;
 DT 11-SEP-1994 (first entry)
 DE Human betal,6-N-acetylglucosaminyltransferase cDNA.
 KW Betal,6-N-acetylglucosaminyl-transferase;
 KW acetylglucosaminotransferase; enzyme; lung carcinoma; tumor;

KW cancer diagnosis; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 156..2093
 FT /*tag= a
 PN EP-585109-A.
 PD 02-MAR-1994.
 PF 24-AUG-1993; 306718.
 PR 24-AUG-1992; JP-245950.
 PR 06-AUG-1993; JP-237118.
 PA (SUNR) SUNTORY LTD.
 PA (TANI/) TANIGUCHI N.
 PI Nishikawa A, Taniguchi N, Yamaguchi N;
 DR WPI; 94-067582/09.
 DR P-PSDB; R48975.
 PT New beta 1,6-N-acetyl glucose aminyl transferase - useful in
 PT sugar synthesis and for identification of cancer inhibitors, also
 PT related DNA, vectors and transformed cells
 PS Claim 2; Page 16; 30pp; English.
 CC DNA encoding this enzyme may be introduced into cells for
 CC expression. The resulting enzyme may be used to introduce,
 CC homogeneously, beta(1-6) branch structures into oligosaccharides.
 CC The enzyme may also be used to screen for specific inhibitors
 CC (possibly cancer metastasis inhibitors).
 SQ Sequence 2095 BP; 602 A; 492 G; 520 T;

Query Match 12.1%; Score 29; DB 1; Length 2095;
 Best Local Similarity 48.2%; Pred. No. 5.9;
 Matches 80; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 57 CCCTGTCTCTCTTGCACCCCTTGGCAGCTCACATGGAACAGGCGCGGTGACTT 126
 Db 487 CCATTGACAAACAGATTGTCACCTTCGACTCCCACTGCCATGGCTGCAAAAGTTA 428

QY 127 TGCACCTGAAGCTGAAGAGTCTTTTCTTGACAAATCTCTATGATGATCCAGCTTCTGG 186
 Db 427 TCTAATAACACAGCAAGGTTTCTTCAGATCATAGCTGTGATGATCCAGCGTATGGC 368

QY 187 AATTGCTTGAAGTCTGCTCTCTCTCCATCTCCCTTCAGGAGC 232
 Db 367 CCAATCCACCATCTCTGTTTCTCTGCGAGTCCCTTGATGATCC 322

RESULT 12
 Q69099/C
 ID Q69099 standard; cDNA; 597 BP.
 AC Q69099;
 DT 26-JAN-1995 (first entry)
 DE ApoE4L protease cDNA.
 KW ApoE4L; protease; enzyme; Alzheimer disease; diagnostic; therapeutic; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 1..595
 FT /*tag= a
 FT /*label= ApoE4L protease
 PN W09413798-A.
 PD 23-JUN-1994.
 PF 16-DEC-1993; E03581.
 PR 04-DEC-1992; CA-085924.
 PR 16-MAR-1993; US-291401.
 PA (BERG) BERGMANN J E.
 PA (PRED/) PREDDIE R E.
 PI Bergmann JE, Preddie RE;
 DR WPI; 94-234212/28.
 DR P-PSDB; R39841.
 PT New proteinase esterase-like proteins - used to develop prods.
 PT for the diagnosis and treatment of Alzheimer's disease and
 PT related diseases
 PS Claim 9; Page 38-39; 72pp; English.
 CC The cDNA encodes ApoE4L, a protease catalyzing the formation of the
 CC abnormal beta/A4 variant of beta-amyloid protein, which is used
 CC to develop an inhibitor for the diagnosis and treatment of Alzheimer

CC disease, Downs syndrome, Parkinson disease, schizophrenia,
CC hyperlipoproteinemia or cardiovascular disease.
SQ Sequence 597 BP; 64 A; 240 C; 197 G; 96 T;

CC hyperlipoproteinemia or cardiovascular disease.
SQ Sequence 597 BP; 64 A; 240 C; 197 G; 96 T;

Query Match 12.1%; Score 29; DB 1; Length 597;
Best Local Similarity 54.1%; Pred. No. 3.8;
Matches 59; Conservative 0; Mismatches 50; Indels

Qy 10 GACTTCAGTGCTCCTCCATCCCAGAGCGCAGTGGCACATATGGGCTCTGGGCTGCCCC 69
|| || || || || || || || || || || || || || || || || || || || || ||
Db 305 GAGCGGGCCCTCAGGCCCATCCGCAGACGCCCTGGGCCCTTGGTAACAGGGCGCGGTG 246

D**b** 305 GAGCGGGGCTCAGGGCCATCCGGGAGCGCCTGGGGCCCCCTGGTGAACAGGGCCCGGTG 246

QY	70	TGTGCTCCTCTTGACCCCTCCTTGGCAGCTCACATGGAACAGGCCGGG	118
Db	245	CGSGCCGCCACTGTGGGCTCCCTGSCCGGCCAGCCGCTACAGGAGCGGG	197

Db 245 CGGGCGGCCACTGTGGCTCCCTGGCCGCCAGCCGCTACAGAGCGG 197

RESULT 13

Q69101/c

ID	Q69101 standard; cDNA; 936 BP.
AC	Q69101.

AC Q69101;
DT 26-JAN-1987

DE ApoE4Lx2 protease cDNA.

PS Claim 13; Fig 6A; 67pp; English.
 CC 2 Open reading frames (T18068 and T18069) identified in the
 CC apolipoprotein E (ApoE) antisense sequence code for proteins
 CC apoE4L (R92113) and apoE4L1 (R92114), respectively. Both ORFs
 CC are transcribed from an upstream regulatory region (T18074).
 CC ApoE4 comprises 2 alleles, one of which causes E4L and E4L1 to
 CC combine to form a hybrid ORF, ApoE4Lx2 (T18070). Elucidation of
 CC the significance of these and other sequences (see also T18065-77)
 CC in the aetiology of Alzheimer's disease (AD) provides means for
 CC diagnosing AD and related diseases, for the design of therapeutic
 CC reagents (e.g. ribozymes or antibodies) and potentially for gene
 CC therapy.
 SQ Sequence 936 BP; 149 A; 317 C; 336 G; 134 T;

Query Match 12.18; Score 29; DB 1; Length 936;
 Best Local Similarity 54.18; Pred. No. 4.4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 10 GACTTCAGTCTCTCCATCCAGAGCGCAGTGGCCACTATGGGCTCTGGGCTGCCCC 69
 || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 305 GAGCGCGCTCAGCGCCATCGCGAGCGCTGGGGCCCTCTGGTGAACAGGGCGCGTG 246
 || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 OY 70 TTGTCTCTCTTGACCTCTCTGGCAGCTCAGTCATGGAGGAGGCGGG 118
 || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 245 CGGGCGGCACCTGTGGCTCTCCCTGGCGCCAGCCGCTACAGAGAGCGGG 197

Search completed: March 20, 2000, 18:37:52
 Job time: 8526 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 21, 2000, 23:48:36 ; Search time 50.59 Seconds
(without alignments)
565.599 Million cell updates/sec

Title: US-09-092-296-1
Perfect score: 239
Sequence: 1 GCCCAGCGGACTTCAGTCT.....CCCTCAGGACGACGCTCA 239

Scoring table: IDENTITY_NUC

Searched: 214294 seqs, 59861208 residues

Database : Issued_Patents_NA.*

Word size : 0

Number of hits that pass the threshold : 428588
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	13.0	3509	3	US-08-817-436A-1	Sequence 1, Appl
2	30.4	12.7	7218	1	US-08-232-463-14	Sequence 14, Appl
3	30.4	12.7	580	6	US-08-627-610-7	Sequence 7, Appl
4	30.4	12.7	580	6	PCT-US95-04636-7	Sequence 7, Appl
5	29	12.1	1157	1	US-07-709-949-1	Sequence 1, Appl
6	29	12.1	2095	2	US-08-405-230-8	Sequence 8, Appl
7	29	12.1	2095	3	US-08-910-990-8	Sequence 8, Appl
8	29	12.1	2259	7	5185254-3	Patent No. 5185254
9	27.6	11.5	5102	2	US-08-494-168-1	Sequence 1, Appl
10	27.6	11.5	3088	2	US-08-418-444A-1	Sequence 1, Appl
11	27.6	11.5	1294	4	US-08-665-647-4	Sequence 4, Appl
12	27.2	11.4	2122	2	US-08-485-449-1	Sequence 1, Appl
13	27.2	11.4	3398	6	PCT-US95-08493-12	Sequence 12, Appl
14	27	11.3	1593	3	US-08-878-989-9	Sequence 9, Appl
15	26.8	11.2	1554	2	US-08-469-486-1	Sequence 1, Appl
16	26.8	11.2	3065	4	US-08-852-153-7	Sequence 7, Appl
17	26.8	11.2	1554	4	US-08-469-658-1	Sequence 1, Appl
18	26.8	11.2	2703	4	US-08-288-508C-1	Sequence 1, Appl
19	26.6	11.1	1241	3	US-08-169-948B-11	Sequence 11, Appl
20	26.6	11.1	1241	3	US-08-448-873-11	Sequence 11, Appl
21	26.6	11.1	1980	4	US-08-766-858A-4	Sequence 4, Appl
22	26.6	11.1	1300	4	US-09-166-203-41	Sequence 41, Appl
23	26.4	11.0	37895	1	US-08-375-709-10	Sequence 10, Appl
24	26.4	11.0	8268	1	US-08-375-709-10	Sequence 10, Appl
25	26.4	11.0	37895	2	US-08-752-929-1	Sequence 1, Appl
26	26.4	11.0	8268	2	US-08-752-929-1	Sequence 1, Appl
27	26	10.9	44377	3	US-08-804-227C-7	Sequence 7, Appl
28	26	10.9	44377	4	US-08-804-198-1	Sequence 1, Appl
29	25.8	10.8	2334	1	US-08-406-070-1	Sequence 1, Appl
30	25.8	10.8	655	3	US-08-812-645-2	Sequence 2, Appl
31	25.6	10.7	944	1	US-08-665-617-1	Sequence 1, Appl
32	25.6	10.7	1215	7	5169941-7	Patent No. 5169941

33	25.4	10.6	252	2	US-08-689-190-1	Sequence 1, Appl
34	25.4	10.6	263	2	US-08-689-190-3	Sequence 3, Appl
35	25.4	10.6	265	2	US-08-689-190-4	Sequence 4, Appl
36	25.4	10.6	234	3	US-08-733-446-19	Sequence 19, Appl
37	25.4	10.6	237	3	US-08-733-446-20	Sequence 20, Appl
38	25.4	10.6	240	3	US-08-733-446-21	Sequence 21, Appl
39	25.4	10.6	243	3	US-08-733-446-22	Sequence 22, Appl
40	25.4	10.6	243	3	US-08-733-446-23	Sequence 23, Appl
41	25.4	10.6	245	3	US-08-733-446-24	Sequence 24, Appl
42	25.4	10.6	247	3	US-08-733-446-25	Sequence 25, Appl
43	25.4	10.6	248	3	US-08-733-446-26	Sequence 26, Appl
44	25.4	10.6	250	3	US-08-733-446-27	Sequence 27, Appl
45	25.4	10.6	251	3	US-08-733-446-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-817-436A-1
; Sequence 1, Application US/08817436A
; Patent No. 5882880
; GENERAL INFORMATION:
; APPLICANT: Canaani, Dan
; TITLE OF INVENTION: Human Checkpoint Gene and Gene for
; TITLE OF INVENTION: Antisense RNA thereof
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wigman, Cohen, Leitner & Myers
; STREET: 900 17th Street, N.W., Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.5 inch, 1.44Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,436A
; FILING DATE: 11-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12445
; FILING DATE: 11-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cohen, Herbert
; REGISTRATION NUMBER: 25,109
; REFERENCE/DOCKET NUMBER: 0744.066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-463-7700
; TELEFAX: 202-463-6915
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3509 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: ESTABLISHED XERODERMA PIGMENTOSUM GM2096-SV3
; IMMEDIATE SOURCE:
; CLONE: RAP-1 cDNA
US-08-817-436A-1

Query Match 13.0%; Score 31; DB 3; Length 3509;
Best Local Similarity 51.0%; Pred. No. 0.45;
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

PCT-US95-04636-7/c
; Sequence 7, Application PC/TUS9504636
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
; NUMBER OF SEQUENCES: 10
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04636
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/346,147
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,511
; FILING DATE: 14-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,812
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,371
; FILING DATE: 14-APR-1994
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91..480
PCT-US95-04636-7

Query Match 12.7% Score 30.4; DB 6; Length 580;
Best Local Similarity 57.3%; Pred. No. 0.31;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 39 CCACTGGCCACTATGGGCTGGGCTGGCCCTTGTCTCTCTTGACCTCTCTGGCAGC 98
Ob 128 CGGTGGCCAGGCGCGGCTCACTGCTGCGGCCCAACATGCTTGTCCCGGCTGTGGC 69
QY 99 TCACATGGAAACAGGCGCGGCTATGACTTTTGCAACTG 134
Db 68 AGAAATGGTCCCTTCGCGGCGGTGAGATTGCTACAG 33

RESULT 5
US-709-949-1
; Sequence 1, Application US/07709949
; Patent No. 5472858
; GENERAL INFORMATION:
; APPLICANT: Attie, Alan D
; APPLICANT: Gretch, Daniel G
; APPLICANT: Sturley, Stephen L
; APPLICANT: Beckage, Nancy E
; TITLE OF INVENTION: Production of Recombinant Proteins in
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Quarles & Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/709,949
; FILING DATE: 19910604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 9629691801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1157 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 62..1015
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 62..119
; PUBLICATION INFORMATION:
; AUTHORS: McLean, J W
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 259
; PAGES: 6498-6504
; DATE: 1984
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1157
US-07-709-949-1

Query Match 12.1% Score 29; DB 1; Length 1157;
Best Local Similarity 54.1%; Pred. No. 1.3;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 10 GACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCACTATGGGTCTGGGTGCCCC 69
Ob 626 GAGCGGCGCTCAGCGCATCCGAGCGCTGGGGCCCTGTGTGAACAGGCGCGGTG 685
QY 70 TTGTCCTCTCTTGACCTCTTGGCTTGGCAGCTCAGATGAACAGGCGCGG 118
Ob 686 CGGGCGGCGCACTGTGGGTCTCCCTGGCGGCGGCGGCTACAGGAGCGGG 734

RESULT 6
US-08-405-230-8/c
; Sequence 8, Application US/08405230
; Patent No. 5707846
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: NISHIKAWA, Atsushi
; APPLICANT: YAMAGUCHI, No. 5707846om1
; TITLE OF INVENTION: NOVEL N-ACETYLGLUCOSAMINYL TRANSFERASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia


```

, TELEPHONE: 415-855-0555
,
, TELEFAX: 415-845-4166
,
, TEL:
,
, INFORMATION FOR SEQ ID NO: 9:
,
,   SEQUENCE CHARACTERISTICS:
,     LENGTH: 1593 base pairs
,     TYPE: nucleic acid
,     STRANDEDNESS: single
,     TOPOLOGY: linear
,   IMMEDIATE SOURCE:
,     LIBRARY: TELYNOT1
,     CLONE: 40194
,   US-08-878-989-9

```

	Query Match	Best Local Similarity	Score 27;	DB 3;	Length 1593;
	Matches 57;	Conservative	Pred. No. 6.8;	Mismatches 0;	Indels 0; Gaps 0;
120	ATGACTTTGCACCTGAAGCTGAAGAGTCTTTTCGTGACAATAATCCTCCTATGATGCCACG	179			
1168	ATGAGTTTGCGAAATTGTGTATAGCTGATTTTTTGTGGAAAATGTTTACTTCAATCAAAG	1109			
180	TTCCCTGGAAATGTCTGAAAAAATGTGCTCCTCCTCCAATCCTCCCTCAAG	227			
1108	GTCCTTTGAGACTCTTGATATTTCTTGCTCTCTCTCTCTGCTTGTCTCTCTGAG	1061			

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; MOLECULE TYPE: CDNA
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..1551
US-08-469-486-1

```

Query Match	11.2%	Score 26.8;	DB 2;	Length 1554;
Best Local Similarity	49.3%;	Pred. No. 7.8;		
Matches 70;	Conservative 0;	Mismatches 72;	Indels 0;	Gaps 0;
QY	46	CCACTATGGGTC	TGGGGCTGCCCTTGCCTCTCTTGACCCCTCTTGCGAGCTCACATG	105
Db	71	CCACATGCGGGGCT	GCTGCATCCTGTGTGTGCAGCACCGCCCTGGGCGGCTCTCTGC	130
QY	106	GAACAGGGCGGGGTATGACCTTGTGCAATGAAGCTGAAGGAGTCTTTTCTTGACAAATTCCT	165	
Db	131	GGCGCGGGGGAGCGT	TCTCTTGGCCCCGGGACACAGGCCCCACGCTGCTTCGAGAGGCC	190
QY	166	CCTATGAGTTCAGCTT	CTCTGGA	187
Db	191	GCAGGGCCAACTCAT	CTTTGGA	212

Search completed: March 22, 2000, 01:22:29
Job time: 5633 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 22, 2000, 04:46:59 ; Search time 802.03 seconds
(without alignments)
1125.125 Million cell updates/sec

Title: US-09-092-296-1
Perfect score: 239
Sequence: 1 GGCACCGGAGCTTCAGTGT.....CCCTTCAGGAGCACCGCTCA 239
Scoring table: IDENTITY_NUC
Searched: 4538634 seqs, 1887831982 residues
Database: EST.*

Word size : 0
Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*

50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
C 1	118	49.4	404	61	AI857998	61	AI857998 wj69b01.x
C 2	74.4	31.1	552	82	A0718761		A0718761 HS_5511_B
C 3	63.6	26.6	328	42	AI136523		AI136523 UI-R-C2p-
C 4	35.6	14.9	509	88	A0838514		A0838514 HS_5011_A
C 5	35	14.6	348	21	F06958		F06958 HSC1QC101 n
C 6	34.4	14.4	238	32	AA376266		AA376266 EST88915
C 7	34.4	14.4	412	38	AA769782		AA769782 ah71b05.s

Seq primer: -40UP from Gibco High quality sequence stop: 395.		Location/Qualifiers	
1. 404		/organism="Homo sapiens"	
/db_xref="taxon:9606"		/clone="IMAGE:2408041"	
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"		/dev_stage="adult"	
/lab_host="DH108 (phage-resistant)"		/notice="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT		97 a 105 c 117 g 84 t 1 others	
ORIGIN			
Query Match		49.4% Score 118; DB 61; Length 404;	
Best Local Similarity		97.7% Pred. No. 7.3e-25;	
Matches 129; Conservative		0; Mismatches 2; Indels 1; Gaps 1;	
QY	109	CAGGCGCGGTATGCTTCACTGCACTGAAGCTGAAGTCTTCTGACAAATTCCTCT 168	
Db	311	CAGGCGCGGTATGCTTCACTGCACTGAAGCTGAAGTCTTCTGACAAATTCCTCT 252	
QY	169	ATGATCTCAGCTTCTGGAATGCTTGAATAA-NTCTGCTCTCTCTCCATCCCTTCAG 227	
Db	251	ATGATCTCAGCTTCTGGAATGCTTGAATAA-NTCTGCTCTCTCTCCATCCCTTCAG 192	
QY	228	GGACGACGCTCA 239	
Db	191	GGACGACGCTCA 180	
RESULT	2		
LOCUS	AQ718761/c	552 bp DNA	GSS 13-JUL-1999
DEFINITION	HS 5511_B2_F09_T7A_RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1087 Col=18 Row=L, genomic survey sequence.		
ACCESSION	AQ718761		
VERSION	AQ718761.1	GI:5468077	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (bases 1 to 552) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		
MEDLINE	99380589		
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallaceu.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jong.med.buffalo.edu). Clones may be purchased from		
ALIGNMENTS			
RESULT	1		
LOCUS	A1857998/c	404 bp mRNA	26-AUG-1999
DEFINITION	wj69b01.x1 NCI_CGAP_Ju19 Homo sapiens cDNA clone IMAGE:2408041 3', mRNA sequence.		
ACCESSION	A1857998		
VERSION	A1857998.1	GI:5511614	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (bases 1 to 404) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3189004. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Sequencing Arrayed by: Greg Lennon, Ph.D. Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jong.med.buffalo.edu). Clones may be purchased from		

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 Or from Research Genetics (Info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1087 row: L column: 18
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 552.

FEATURES
 source
 Location/Qualifiers
 1..552

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-1087 Col-18 Row=L"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Genomic sequence of BAC ends"
 BASE COUNT 141 a 153 c 124 g 117 t 17 others
 ORIGIN

Query Match 31.1%; Score 74.4; DB 82; Length 552;
 Best Local Similarity 87.1%; Pred. No. 6.3e-12;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 107 AACAGGCGGGTATGACTTCACTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTC 166
 |||||
 Db 474 AGCATGGCGAGGATGACTTCGACCTGAAGCTGAAGGAGTCTATTAGACAAATTCGTC 415
 |||||

Oy 167 CTATGATCCAGCTTCTGTAATTCCTTGAAAA 199
 |||||

Db 414 TTATGATCCAGCTTCTGTAATTCCTTGAAAA 382
 |||||

RESULT 3
 AIL36523/c
 LOCUS
 DEFINITION UI-R-C2p-ng-e-02-0-UI-sl UI-R-C2p Rattus norvegicus cDNA clone
 UI-R-C2p-ng-e-02-0-UI 3', mRNA sequence.
 AIL36523
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 328)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150222.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult lung library. cDNA Library Preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..328

/organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C2p-ng-e-02-0-UI"
 /clone_lib="UI-R-C2p"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"

FEATURES
 source

/note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p
 library is a subtracted library derived from the UI-R-C1
 library, which is a subtracted library derived from the
 UI-R-C0 library. The UI-R-C0 library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dT track which allows
 identification of the library of origin of a clone within
 the mixture. The subtracted library (UI-R-C2p) was
 constructed as follows: PCR amplified cDNA inserts from
 UI-R-C1 clones from which 3' ESTs had been derived was
 used as a driver in a hybridization with the UI-R-C1
 library in the form of single-stranded circles. The
 remaining single-stranded circles (subtracted library) was
 purified by hydroxyapatite column chromatography,
 converted to double-stranded circles and electroporated
 into DH10B bacteria (Life Technologies) to generate the
 UI-R-C2p library. This procedure has been previously
 described (Bonaldo, Lennon and Soares, Genome Research 6:
 791-806, 1996)"

BASE COUNT 62 a 77 c 98 g 91 t
 ORIGIN

Query Match 26.6%; Score 63.6; DB 42; Length 328;
 Best Local Similarity 68.7%; Pred. No. 7.9e-09;
 Matches 101; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

Oy 94 GCAGCTCACATGACAGCGCGGGTATGACTTTCGAACCTGAAGCTGAAGGAGTCTTTC 153
 |||||
 Db 327 GCAGCTCACATGACAGCGCGGCAACAAAGGTGACTTTCGAAGTAAACCTGACCGGAGCATCTC 268
 |||||

Oy 154 TGCAAAATTCCTCTATGATGATCCAGCTTCTGTAATTCCTTGAAAA-NTCTGCCCTCTCC 212
 |||||
 Db 267 AGGCAAGACCTCCCAAGACTCGGCTTCTGACATGCTCCAAAAGATCTGCTCTCTCC 208
 |||||

Oy 213 TCCAICTCTCCCTTCAGGACAGCGTCA 239
 |||||

Db 207 TCCACCTCTCAGCGGGGCCAATGTCA 181
 |||||

RESULT 4
 A0838514
 LOCUS
 DEFINITION HS.501L.A2.D10.SP6E RPCI-11 Human Male BAC Library Homo sapiens
 genomic Clone Plate-587 Col-20 Row-G, genomic survey sequence.
 A0838514
 ACCESSION
 VERSION A0838514.1 GI:5808388
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 509)
 AUTHORS Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University Of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (<http://www.htsc.washington.edu>) Plate: 587 row: G column: 20

Seq primer: SP6
Class: BAC ends
High quality sequence stop: 509.
Location/Qualifiers
1. .509

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate:587 Col=20 Row=G"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Genomic sequence of BAC ends"

BASE COUNT 120 a 127 c 106 g 153 t 3 others
ORIGIN

Query Match 14.9%; Score 35.6; DB 88; Length 509;
Best Local Similarity 47.3%; Pred. No. 1.8;
Matches 104; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 8 GGACTTCAGTCTCTCCATCCAGAGCGCAGTGGGCATATGCGGTCTGGCTGCC 67
Db 80 GTGACTCTTTTCTCTCACCCCATGCTCTCAGGAGGAGCAGTGGCTTTCTTC 139
Qy 58 CTTGTCTCTCTCTTGTACCTCTCTGGCAGCTCACATGGAACAGCGCGGTATGACTTT 127
Db 140 CCTCTGTCTATATGCTCTCTCCCTGCAAGCTGCCAACCACTGGTCTTACAGCTA 199
Qy 128 GCACTGAGCTGAAGAGTCTTTCTGCAAAATCTCTATGAGTCCAGCTCTCTCGA 187
Db 200 GTAGTGGAGATGCCAAGTGTATGTATTAATGAGCCTCTCTCTGAGTCCAGGOC 259
Qy 188 ATTGCTTGAAATCTCGCTCTCTCTCCATCTCCCTTCAG 227
Db 260 ATGTAAGTTACAGTGGCACCAGCAGTGGCCCTCTGGAG 299

RESULT 5

LOCUS F06958 348 bp mRNA EST 20-FEB-1995
DEFINITION HSC1Q101 normalized infant brain cDNA Homo sapiens cDNA clone c-lq10, mRNA sequence.

ACCESSION F06958

VERSION F06958.1 GI:672595

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 348)

Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houllgate, R., Juneau, M.N., Lamy, B., Lorenz, F., Michell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y., Sebastiani-Ksaktchis, C. and Tessier, A.

IMAGE: molecular integration of the analysis of the human genome and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

9527534

On Sep 21, 1992 this sequence version replaced gi:278976.

Contact: Genethon

Genethon-Genethon

Genethon Centre de recherche sur le Genome Humain

1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE

Tel: 3316942800

Fax: 33160778698

Email: genexpress@genethon.fr

Single read.

Genexpress_library_id: C; Genexpress_sequence_id: ylc-lq10

Insert Length: 639 Std Error: 0.00

Seq primer: (-21)M13_universsl
High quality sequence stop: 150.
Location/Qualifiers
1. .348

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-lq10"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"

/note="Organ: brain; Vector: lafmid BA; Site 1: HindIII; Site 2: NotI; sex: female; dev_stage=3 months old; isolate-muscular atrophy patient; tissue_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B. Soares, Psychiatry Dept. Columbia University, USA. Normalization_method: Bento Soares, P.N.A.S in press"

BASE COUNT 84 a 91 c 90 g 81 t 2 others
ORIGIN

Query Match 14.68; Score 35; DB 21; Length 348;

Best Local Similarity 63.1%; Pred. No. 2.3; Mismatches 31; Indels 0; Gaps 0;
Matches 53; Conservative 0;

Qy 18 TGTCTCTCCATCCAGGAGCGCAGTGGGTCTGGGTCTGGCTTCCTCTCCTC 77
Db 5 TGTGCTCTCTCCAGGATCCCTTGTGAGTATGTTTCAGGNTGCACCACCAC 64
Qy 78 CTCCTGACCTCTCTGGCAGCTCA 101
Db 65 CTCATAGATACCTTCAGGCAACACA 88

RESULT 6

LOCUS AA376266 238 bp mRNA EST 21-APR-1997
DEFINITION EST88915 HSC172 cells II Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION AA376266

VERSION AA376266.1 GI:2028809

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 238)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Sprague, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dmke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hung, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

On Apr 14, 1993 this sequence version replaced gi:693135.

Other ESTs: THC191210

Contact: Kerlavage, AR

Bioinformatics

TITLE

JOURNAL

MEDLINE

COMMENT

THEORY

```
lab_host="DH1"
```

BASE COUNT 128 a 92 c 130 g 108 t
ORIGIN

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared; and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

Query Match 14.4%; Score 34.4; DB 49; Length 458;
Best Local Similarity 63.1%; Pred. No. 3.8;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

RESULT 9
LOCUS AI803529/470 bp mRNA EST 06-JUL-1999
DEFINITION tc42hl1.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:2067333 3', mRNA sequence.

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

QY 18 TGTCTCTCTCATCCAGGAGCGAGTGGCCACTATGGGGTCTGGGCTGGCCCTTGTCTCTC 77
Db 316 TGTGGCTCTCTCCAGGATCCCTTGGTGAGTATGGTGTCAGGATGCCACCACCA 257

QY 78 CTCTTGACCTCTTGGCAGCTCA 101
Db 256 CTCTAGATACCTTCAGGCAACACA 233

BASE COUNT	CONTROL		
ORIGIN	173 a	134 c	
	120 g	109 t	1 others

Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press.
BASE COUNT 57 a 116 c 100 g 64 t 2 others
ORIGIN

Query Match 13.5%; Score 32.2; DB 21; Length 339;
Best Local Similarity 58.5%; Pred. No. 15; Mismatches 39; Indels 0; Gaps 0;
Matches 55; Conservative 0;

QY 23 CTTCCATCCAGGAGCGGACATGATGGGTGTGGGTCGCCCTTGTCTCTCTT 82
DB 61 CCCACTGGCGTGTCTCTCCGCCCTCCGGAAGCTTGGATGCCCTNACACCCCTCTT 120

QY 83 GACCTTCCTGTGGAGCTCACATGGACAGGGCGG 116
DB 121 GATCTTCCCTGTGATGTACCTGGACCCCTGCTG 154

RESULT 15
278408 742 bp mRNA EST 28-JUL-1999
LOCUS HS278408 Human fetal brain S. Meier-Ewert Homo sapiens cDNA clone
DEFINITION 3.99 (CEPH), mRNA sequence.
ACCESSION 278408
VERSION 278408.1 GI:1495181
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 742)
Neri, C., Albanese, V., Lebre, A. S., Holbert, S., Saada, C.,
Bougueret, L., Meier-Ewert, S., Legall, I., Millasseau, P., Bui, H.,
Giudicelli, C., Massart, C., Guillon, S., Gervy, P., Poullier, E.,
Rigault, P., Weissenbach, J., Lennon, G., Chumakov, I., Dausset, J.,
Lehrach, H., Cohen, D. and Cann, H. M.
Survey of CAG/CTG repeats in human cDNAs representing new genes:
candidates for inherited neurological disorders
Hum. Mol. Genet. 5 (7), 1001-1009 (1996)
95414310
On Apr 14, 1993 this sequence version replaced gi:837709.
Contact: Neri C.
Fondation Jean Dausset - CEPH
27 Rue Juliette Dodu, 75010 Paris, France
ICRF clone ID ICRFp507L04199.
LOCATION/Qualifiers
1. .742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="3.99 (CEPH)"
/clone_lib="Human fetal brain S. Meier-Ewert"
/tissue_type="brain"
/dev_stage="fetus"
/note="cDNA library of S. Meier-Ewert, Max Planck
Inst.f.Mol.Genetics, Berlin, FRG"

BASE COUNT 145 a 232 c 215 g 131 t 19 others
ORIGIN

Query Match 13.5%; Score 32.2; DB 27; Length 742;
Best Local Similarity 53.8%; Pred. No. 21;
Matches 64; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 19 GTCTCTCCATCCAGGAGCGGACATGATGGGTCTGGGTCGCCCTTGTCTCTCC 78
DB 198 GGCTACTCTAGGGATAGCCCGCGTCCCTCATCTCCGTGTCAGTGTCCCTGACCCCCA 257

QY 79 TCTTGCACCTCTCTTGGCAGCTCACATGGAAACAGGCGCGGGATGACTTTGCAACTGAAG 137
DB 258 TCTACTCTCTTCTGGGGACTTCTCAGTCACAGGGCTGGCGCTGGACTTTCANCTGNAG 316

WILEY

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 20 12:32:52 2000; MasPar time 5.78 Seconds
319.798 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-15
Description: (1-78) from US0902296.pep
Perfect Score: 558

Sequence: 1 MSGSLPIVLLTLGSSHGT.....SGTSTLHARSQHVVVNCNT 78

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: a-geneseq35
1-geneseqp

Statistics: Mean 27.552; Variance 131.690; scale 0.209

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	558	100.0	78	1 W88498	Human stomach carcinom	3.43e-38
2	86	15.4	46	1 W70327	Secreted protein FB/8_	4.56e-01
3	86	15.4	47	1 W73408	Human secreted protein	4.56e-01
4	85	15.2	325	1 R48730	G-protein coupled bovi	5.33e-01
5	85	15.2	325	1 W02702	G-protein coupled bovi	5.33e-01
6	84	15.1	154	1 W98259	H. pylori GHPO 895 pro	6.24e-01
7	84	15.1	795	1 W97842	Human P2Y11 receptor.	6.24e-01
8	83	14.9	194	1 R47338	Peptide fragment of te	7.29e-01
9	83	14.9	411	1 W57046	Mouse apoptosis induc	7.29e-01
10	82	14.7	110	1 R26954	Human T lymphocyte rec	8.52e+01
11	82	14.7	169	1 R21674	Human mitochondrial ei	8.52e+01
12	81	14.5	401	1 W59924	Human 7-transmembrane	9.94e-01
13	81	14.5	1253	1 W10038	Mad binding protein, m	9.94e-01
14	81	14.5	1261	1 W10040	Mad binding protein, m	9.94e-01
15	80	14.3	398	1 R15138	Human serotinin ID rec	1.16e-02
16	80	14.3	713	1 R10052	Cyclomaltodextrin gluc	1.16e-02
17	80	14.3	713	1 R06110	Sequence of cyclomalto	1.16e-02
18	80	14.3	945	1 R60600	Human NMDAR2 receptor	1.16e-02
19	80	14.3	1214	1 R60605	Human NMDAR2 receptor	1.16e-02
20	80	14.3	1214	1 W87509	Human N-methyl-D-aspar	1.16e-02
21	80	14.3	1219	1 W87507	Human N-methyl-D-aspar	1.16e-02
22	80	14.3	1219	1 R65063	Human NMDAR2 receptor	1.16e-02
23	80	14.3	1231	1 R65062	Human NMDAR2 receptor	1.16e-02

24	80	14.3	1231	1 W87506	Human N-methyl-D-aspar	1.16e-02
25	80	14.3	1236	1 R66037	Human N-methyl-D-aspar	1.16e-02
26	80	14.3	1236	1 W85574	Human N-methyl-D-aspar	1.16e-02
27	80	14.3	1239	1 W87508	Human N-methyl-D-aspar	1.16e-02
28	80	14.3	1239	1 R65054	Human NMDAR2 receptor	1.16e-02
29	80	14.3	1244	1 R65054	Human NMDAR2 receptor	1.16e-02
30	80	14.3	1244	1 W87505	Human N-methyl-D-aspar	1.16e-02
31	79	14.2	620	1 W14993	Human c-Fos induced gr	1.35e-02
32	79	14.2	652	1 W49879	Amino acid sequence of	1.35e-02
33	79	14.2	907	1 W93889	Human HG38 protein.	1.35e-02
34	79	14.2	969	1 W25170	Human insulinoma-assoc	1.35e-02
35	79	14.2	986	1 W25171	Type I diabetes-associ	1.35e-02
36	79	14.2	1015	1 W18092	Human protein tyrosine	1.35e-02
37	79	14.2	1015	1 W35345	Human secreted protein	1.35e-02
38	78	14.0	120	1 W67828	Kaposi's sarcoma assoc	1.58e-02
39	78	14.0	120	1 R97838	Kaposi's sarcoma assoc	1.58e-02
40	78	14.0	127	1 R93614	C-Delta-1 polypeptide.	1.58e-02
41	78	14.0	727	1 W11719	C-Delta-1 polypeptide.	1.58e-02
42	78	14.0	740	1 W00876	Chlamydia pneumoniae s	1.58e-02
43	78	14.0	922	1 W88419	Glutamic acid receptor	1.58e-02
44	78	14.0	1239	1 R45945	H. pylori GHPO 343 pro	1.83e-02
45	77	13.8	246	1 W98638		

ALIGNMENTS

RESULT 1
ID W88498 standard; Protein; 78 AA.
AC W88498;
DT 30-MAR-1999 (first entry)
DE Human stomach carcinoma HP10408-encoded transmembrane protein.
KW Transmembrane protein; HP10408; human; stomach cancer.
OS Homo sapiens.
FN W09855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI: 99-045730/04.
DR N-PSDB: W84366.
PT New human proteins containing transmembrane domains and their
PT encoding sequences - useful in the preparation of antibodies and
PT large-scale protein production, gene diagnosis, and gene therapy
PS Claim 1; Page 135; 178pp; English.
CC This is the amino acid sequence of a novel transmembrane protein
CC encoded by human stomach cancer cDNA clone HP10408 (see W84366).
CC The encoded protein has a putative signal sequence and a putative
CC internal transmembrane domain. The invention provides nucleotide
CC sequences (see W84359-76) coding for 18 transmembrane proteins
CC (see W88491-509), vectors containing such polynucleotides, and
CC eukaryotic cells containing the vectors. The proteins can be
CC used as antigens or as compositions in the preparation of
CC antibodies against the proteins. The polynucleotides can be used
CC as probes for gene diagnosis, and as gene sources for gene therapy
CC and large-scale production of proteins encoded by the cDNA. The
CC host cells are used for the detection of ligands corresponding to
CC the expressed proteins, and the screening of low mol.wt. medicines.
SQ Sequence 78 AA;

Query Match 100.0%; Score 558; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 3.43e-38;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MSGSLPIVLLTLGSSHGTGFMTLQKLKSFNTNSYESSFLELKLCLLHPSG 60

QY 1 MSGSLPIVLLTLGSSHGTGFMTLQKLKSFNTNSYESSFLELKLCLLHPSG 60

Db 61 TSVTLHARSQHVVVNCNT 78

QY 61 TSVTLHARSQHVVVNCNT 78

```

RESULT      2
ID   W70327 standard; Protein; 46 AA.
AC   W70327;
DT   21-DEC-1998 (first entry)
DE   Secreted protein FB78.1.
KW   Secreted protein; DU123.1; human.
OS   Homo sapiens.
FH   Key      Location/Qualifiers
FT   Peptide  8..20
FT   /note= "predicted leader/signal sequence, or
FT           transmembrane domain"
FT   W09838209-A2.
PD   03-SEP-1998.
PR   25-FEB-1998: U03697.
PR   24-FEB-1998: US-028724.
PR   26-FEB-1997: US-805819.
PA   (GENY ) GENETICS INST INC.
PI   Agostino M, Jacobs K, LeVallie ER, McCoy JM, Werberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-481139/41.
DR   N-PSDB; V33199.
PT   New isolated polynucleotide(s) and encoded polypeptide(a) -
PT   obtained from human foetal kidney, adult colon, adult brain, foetal
PT   brain and placenta cDNA libraries.
PS   Claim 36; Page 83; 103pp; English.
CC   This is the amino acid sequence of novel human secreted protein
CC   FB78.1, as deduced from a full-length cDNA clone (see V33199)
CC   obtained from a human adult placenta cDNA library. Database
CC   searching revealed some similarity between FB78.1 and some known
CC   sequences. The invention provides new isolated polynucleotides
CC   (see V33199-99), from human foetal kidney, adult colon, adult brain,
CC   foetal brain and placenta cDNA libraries, that code for secreted
CC   proteins (see W70319-27). The clones can be used for recombinant
CC   production of the polypeptides, which may have activities such as
CC   e.g. nutritional activity, cytokine and cell proliferation or
CC   differentiation activity, immunostimulant or immunosuppressive,
CC   haematopoietic regulating activity, tissue growth activity, activin
CC   or inhibin activity, chemotactic or chemokinetic activity,
CC   haemostatic and thrombolytic activity, receptor/ligand activity,
CC   antiinflammatory activity, cadherin/tumour invasion suppressor
CC   activity, tumour inhibition activity, or other activities.
SQ   Sequence 46 AA;

Query Match      15.4%; Score 86; DB 1; Length 46;
Best Local Similarity 56.5%; Pred. No. 4.56e+01;
Matches 13; Conservative 6; Mismatches 3; Indels 1; Gaps 1;

Db   6 GAALPILLLLALRGTFHGARGP 28
Oy   2 GSGLP-LVLLTLGSSHGTCGPG 23
      |::|| |::|| |::|| |::||
      1::|| |::|| |::|| |::||

RESULT      3
ID   W73408 standard; Protein; 47 AA.
AC   W73408;
DT   19-FEB-1999 (first entry)
DE   Human secreted protein encoded by Gene No. 12.
KW   Secreted protein; human; protein therapy; gene therapy; blood disorder;
KW   pathological condition; diagnosis; cancer; neurological disorder;
KW   developmental abnormality; foetal deficiency; leukaemia; hepatic disease;
KW   immune system abnormality; Alzheimer's disease; cognitive disorder;
KW   schizophrenia; prostate disease; autoimmune disorder; AIDS.
OS   Homo sapiens.
FH   Key      Location/Qualifiers
FT   Misc_difference 47
FT   /note= "unspecified amino acid"
FT   W09854206-A1.
PD   03-DEC-1998.
PR   28-MAY-1998: U10868.
PR   29-AUG-1997: US-058296.
PR   30-MAY-1997: US-044039.
PR   30-MAY-1997: US-048093.

```

```

PR   30-MAY-1997: US-048101.
PR   30-MAY-1997: US-048190.
PR   30-MAY-1997: US-048356.
PR   30-MAY-1997: US-050935.
PR   29-AUG-1997: US-056250.
PR   29-AUG-1997: US-056293.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Carter KC, Dillon PJ, Endress GA, Feng P, Ni J,
PI   Rosen CA, Ruben SW, Yu G;
DR   WPI; 99-070209/06.
DR   N-PSDB; V08822.
PT   New isolated human genes - useful for diagnosis and treatment of,
PT   e.g. cancers, neurological disorders, immune diseases, developmental
PT   disorders or blood disorders
PS   Claim 11; Page 152-153; 188pp; English.
CC   This sequence is encoded by a cDNA of the invention, designated
CC   Gene No. 12. This sequence represents a human secreted protein, and is
CC   expressed in activated neutrophils, endothelial cells, T-cells and
CC   to a lesser extent in brain and liver.
CC   The DNA sequences of the invention and their corresponding secreted
CC   polypeptides are useful for preventing, treating or ameliorating medical
CC   conditions, e.g. by protein or gene therapy. Also pathological conditions
CC   can be diagnosed by determining the amount of the new polypeptides in a
CC   sample or by determining the presence of mutations in the DNA sequences.
CC   Specific uses are described for each of the DNA sequences and the encoded
CC   proteins, based on which tissues they are most highly expressed in, and
CC   include developing products for the diagnosis or treatment of cancer,
CC   tumours, neurological disorders, developmental abnormalities and foetal
CC   deficiencies, blood disorders, leukaemias, diseases of the immune system
CC   (including allergies or asthma), hepatic disease, Alzheimer's and
CC   cognitive disorders, schizophrenia, prostate disease, autoimmune
CC   disorders and AIDS. The polypeptides are also useful for identifying
CC   their binding partners.
SQ   Sequence 47 AA;

Query Match      15.4%; Score 86; DB 1; Length 47;
Best Local Similarity 56.5%; Pred. No. 4.56e+01;
Matches 13; Conservative 6; Mismatches 3; Indels 1; Gaps 1;

Db   6 GAALPILLLLALRGTFHGARGP 28
Oy   2 GSGLP-LVLLTLGSSHGTCGPG 23
      |::|| |::|| |::|| |::||
      1::|| |::|| |::|| |::||

RESULT      4
ID   R48730 standard; Protein; 325 AA.
AC   R48730;
DT   06-JUN-1996 (first entry)
DE   G-protein coupled bovine adrenal angiotensin II type-1 receptor protein.
DE   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   psychotic disorder; schizophrenia; dopamine; CAMP; adenosine; thrombin;
KW   muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
KW   rhodopsin; opsin; odorant; cytomegalovirus.
OS   Bos taurus.
PN   W09405695-A1.
PD   17-MAR-1994.
PR   09-SEP-1993; U08528.
PR   10-SEP-1992; US-943236.
PA   (UNYV ) UNIV NEW YORK STATE.
PI   Murphy RB, Schuster DI;
DR   WPI; 94-101120/12.
PT   Polypeptides of G-coupled receptor proteins (GPRs) - useful for
PT   binding GPR ligands or modulating GPR binding
PS   Disclosure; Page 106-107; 160pp; English.
CC   Proteins R48685-R48758 represent a range of G-protein coupled receptor
CC   proteins selected from CAMP, adenosine, muscarinic acetylcholine, opsin,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other G-protein coupled receptors. The
CC   receptor proteins were used to design polypeptides, pref. based on the
CC   transmembrane domains, for use in G-protein coupled receptor ligand
CC   binding assays. The polypeptide fragments retain biological activity
CC   such as binding a GPR ligand or modulating GPR ligand binding to a GPR
CC   (see R48759-R48758, R50569-R50807 and R89189-R89195 for examples of

```


CC polypeptide fragments). The polypeptide fragments can be used in
 CC compositions for treating subjects suffering from a pathology related to
 CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
 SQ Sequence 325 AA;

Query Match 15.2%; Score 85; DB 1; Length 325;
 Best Local Similarity 44.4%; Pred. No. 5.33e+01;
 Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db 27 YMKLKTASVFLNLALADLCFLTLTP 53
 QY 34 FLTNSSYESSFL-EL-LEKLKLLHLHP 58

RESULT 5
 ID W02702 standard; peptide; 325 AA.
 DT 13-NOV-1996 (first entry)

DE G-protein coupled bovine adrenal angiotensin II type-1 receptor.
 KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
 KW muscarinic acetylcholine; bombesin; bombesin; endocrine; rhodopsin;
 KW odorant; cytomagalovirus; serotonergic.
 OS Bos taurus.

PN US508384-A.
 PD 16-APR-1996.

PF 10-SEP-1992; 943236.

PR 10-SEP-1992; US-943236.

PR 09-SEP-1993; US-118270.

PA (UYNV) UNIV NEW YORK STATE.

PI Murphy EB, Schuster DI;

DR WPI; 36-208785/21.

PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.

PT for treating schizophrenia

PS Disclosure; Column 129-132; 184pp; English.

CC Proteins W02657-W02730 represent a range of G-protein coupled receptor
 CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
 CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomagaloviral and other GPR proteins. The receptor proteins
 CC were used to design polypeptides, pref. based on the transmembrane
 CC domains, for use in G-protein coupled receptor ligand binding assays.

CC The polypeptide fragments retain biological activity such as binding a
 CC GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
 CC for examples of polypeptide fragments). The polypeptide fragments can
 CC be used in compositions for treating subjects suffering from a pathology
 CC related to a GPR abnormality e.g. a psychotic disorder such as
 CC schizophrenia.
 SQ Sequence 325 AA;

Query Match 15.2%; Score 85; DB 1; Length 325;
 Best Local Similarity 44.4%; Pred. No. 5.33e+01;
 Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db 27 YMKLKTASVFLNLALADLCFLTLTP 53
 QY 34 FLTNSSYESSFL-EL-LEKLKLLHLHP 58

RESULT 6
 ID W98259 standard; Protein; 154 AA.

AC W98259;

DT 31-MAR-1999 (first entry)

DE H. pylori GHPO 895 protein.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

KW peptic ulcer disease.

OS Helicobacter pylori.

PN W09843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; U06371.

PR 29-JUL-1997; US-902615.

PR 01-APR-1997; US-833457.

PR 24-JUN-1997; US-881227.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Kleantous H, Miller C, Omen NP, Tomb J;
 DR WPI; 98-542293/46.
 DR N-PSDB; X13978.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 8; Page 250-251; 2054pp; English.
 CC This sequence represents a Helicobacter pylori GHPO protein of the
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SQ Sequence 154 AA;

Query Match 15.1%; Score 84; DB 1; Length 154;
 Best Local Similarity 46.9%; Pred. No. 6.24e+01;
 Matches 15; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Db 74 LGGFGFGGLGLGLGFWFVTFSSFFSSFL 105
 QY 14 LGSSTGTGPGMTLQKLKESFLTNSYESSFL 45

RESULT 7
 ID W97842 standard; Protein; 795 AA.
 AC W97842;
 DT 07-JUN-1999 (first entry)
 DE Human P2Y11 receptor.
 KW P2Y11; G protein coupled receptor; human; infection; neutropaenia;
 KW agranulocytosis; cancer; leukaemia; diagnosis; therapy.
 OS Homo sapiens.
 FH Key
 FT Modified_site 474 Location/Qualifiers
 FT /note= "putative protein kinase C phosphorylation
 FT site"
 FT Modified_site 600 /note= "N-glycosylation"
 FT Modified_site 794 /note= "calmodulin-dependent protein kinase
 FT phosphorylation site".
 FT W09902675-A1.
 PN 21-JAN-1999.
 PD 09-JUL-1998; BE0108.
 PR 09-JUL-1997; EP-870101.
 PA (EURO-) EUROSREEN SA.
 PI Boeynaems J, Communi D;
 DR WPI; 99-120876/10.
 DR N-PSDB; X07369.

PT New G protein-coupled receptor - useful for diagnosis, treatment and
 PT prevention of neutropaenia, agranulocytosis, infection and cancer
 PS Claim 2; Fig 1; 46pp; English.
 CC This polypeptide comprises a novel human G protein coupled receptor,
 CC termed P2Y11 that has selective affinity for ATP. The amino acid
 CC sequence was deduced from genomic DNA clones (see X07369). The
 CC invention also provides vectors, transformed cells, anti-P2Y11
 CC antibodies, nucleic acid probes, pharmaceutical compositions
 CC comprising such products and transgenic animals. Antisense
 CC nucleotides (claimed) that hybridise to mRNA are used to decrease
 CC activity of P2Y11, while specific antibodies are used to block
 CC binding of P2Y11 to its ligand. Probes are used in hybridisation
 CC assays to detect expression of P2Y11 at the RNA level, while
 CC antibodies are used similarly at the protein level in standard
 CC immunoassays, particularly for diagnosis of leukaemia. The
 CC transgenic animals are used to determine the effects of varying
 CC levels of P2Y11 expression. These animals, and host cells, are
 CC used in drug screening methods to identify (antagonists that are
 CC potentially useful for treatment or prevention of disorders
 CC associated with excessive or inadequate receptor activity,
 CC specifically neutropaenia, agranulocytosis, infections and cancer.
 CC Host cells are also used to produce recombinant P2Y11.

Query Match 14.5%; Score 81; DB 1; Length 401;
Best Local Similarity 38.1%; Pred. No. 9.94e+01;
Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

RESULT 13
ID W10038 standard; Protein; 1253 AA.

AC	W10038; 1998	(first entry)
AD	Mad binding protein, msfA.	
DE	Kw murine; msfA; mammalian homologue; Saccharomyces cerevisiae; repressor;	
KW	Sin3; Mad; Max; Mad complex; mSin; Mad; Max complex; Myc; promoter;	
KW	basic helix-loop-helix zipper protein; compete; DNA-binding;	
KW	Myc; Max complex; activate; transcription; gene regulation.	
OS	Mus musculus.	

Misc_difference	10	/label=	unknown
FT		/note=	"encoded by TAG"
FT		1238	
FT		/label=	unknown
FT		/note=	"encoded by TAG"
US5624818.A.			
PD	29-APR-1997		
PN	01-JUN-1994;	252966	
PF	01-JUN-1994;	US-252966	
PR	19-SEP-1991;	US-756195	
PR	23-JUN-1994;	US-903710	
PR	01-APR-1994;	US-222638	
PR	(HUTC-) HUTCHINSON CANCER RES CENT FRED.		
PI	AYER DE, Eisenman RN;		
PI	WPI: 97-259216/23.		
DR	N-PSDB; T70126.		
DR	ms1n nucleic acids encoding recombinant polypeptide(s) that		
PT	associate with Mad polypeptide - are possible homologues of S.		
PT	cerevisiae general repressor protein Sin3		
PS	Example 12; Fg 23A-C; llbp; English.		
CC	This sequence represents the murine protein, designated ms1nA, which may		
CC	be a mammalian homologue of the saccharomyces cerevisiae general		
CC	repressor protein Sin3. The ms1n protein associates with a Mad		
CC	polypeptide to form a ms1n:Mad complex, which preferably associates		
CC	with a Max polypeptide to form a ms1n:Mad:Max complex which binds to a		
CC	nucleotide sequence comprising CACGGT. Mad is a basic helix-loop-helix		
CC	(bHLH) zipper protein which can compete with Myc by forming sequence-		
CC	specific DNA-binding heterocomplexes with Max. Mad:Max complexes repress		
CC	while Myc:Max complexes activate, transcription from promoters containing		
CC	proximal CACGGT binding sites for these proteins. Expression of Mad is		
CC	closely linked to differentiation in at least two distinct cell lineages.		
CC	The switch from Myc:Max to Mad:Max complexes may reflect the repression		
CC	of transcription of Myc regulated genes by Mad. The DNA, vectors and host		
CC	cells of the invention are useful for the recombinant production of ms1n		
CC	proteins useful in elucidation of Mad repressor functions.		
SQ	Sequence 1253 AA;		

KW	5-HT (LD); Parkinson's Disease; migraine; anxiety; eating disorder;
KW	G-protein; 5-hydroxytryptamine.
OS	Homo sapiens.
FH	Key
FT	Location/Qualifiers
FT	1..49
FT	/label= extracellular
FT	domain
FT	50..73
FT	/label= transmembrane-1
FT	domain
FT	74..86
FT	/label= cytoplasmic
FT	domain
FT	87..110
FT	/label= transmembrane-2
FT	domain
FT	111..121
FT	/label= extracellular
FT	domain
FT	122..145
FT	/label= transmembrane-3
FT	domain
FT	146..164
FT	/label= cytoplasmic
FT	domain
FT	165..188
FT	/label= transmembrane-4
FT	domain
FT	189..205
FT	/label= extracellular
FT	domain
FT	206..229
FT	/label= transmembrane-5
FT	domain
FT	230..315
FT	/label= cytoplasmic_loop
FT	domain
FT	316..339
FT	/label= transmembrane-6
FT	domain
FT	340..348
FT	/label= extracellular
FT	domain
FT	349..372
FT	/label= transmembrane-7
FT	region
FT	373..398
FT	/label= cytoplasmic_tail
PN	WO911714-A.
PD	14-NOV-1991.
PR	08-MAY-1991; U03200.
PR	08-MAY-1990; US-520716.
PA	(NEUR-) NEUROGENETIC CORP.
PI	Weinshank RL, Branchek T, Hartig PR;
PI	WPI: 91-353715/48.
DR	N-PSDB: O14836.
DR	Nucleic acid encoding 5HT-LD receptors and their antibodies -
PT	used to treat and diagnose conditions caused by abnormal 5HT-LD
PT	receptor expression e.g. dementia
PT	Claim 7; Fig 4; 90pp; English.
CC	CS There are two subtypes of 5HT(LD) receptors encoded by gene 8-30-84
CC	and gene 11, respectively. A full-length clone corresponding to
CC	gene 11 was isolated from a human placental library. The
CC	amino acid sequence deduced from the nucleotide sequence predicts a
CC	mol.wt. 44,333 protein. Comparison of this protein sequence with
CC	previously characterised neurotransmitter receptors suggests that
CC	it is a new member of the G protein-coupled receptor family.
CC	See Q14836 for gene 11.
SC	Sequence 398 AA;
SQ	Query Match 14.3%; Score 80; DB 1; Length 398;
	Best Local Similarity 23.6%; Pred.No.1.16e+02;
	Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;
DB	52 LVMLALITATLTSNAFVIATYVRFKRLTPANKVLASIDVTDLIVSLVIPISMTVV 111
OY	::: : : : : :: : : :: : : : : : : :
	7 LVLLTLGSHGGTGPMGLQLKLKESFL-TNSSVESFFLEELKLCILLHLPGSTGYL 65
DB	112 TDRWLISQWCD 123
OY	::: ::
	66 HHANSOHHWCN 77

Search completed: Mon Mar 20 12:33:02 2000
Job time : 10 secs.


```
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 325 AA; 3728 MW; 592694 CN;

Query Match      15.2%; Score 85; DB 1; Length 325;
Best Local Similarity 44.4%; Pred. No. 2.28e+01;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db    27 YMKLKYASYVFLNALADLCFLTLTP 53
      ::|:|||::|||::|::|::|::|::|
QY    34 FLTNSSYSESL-EL-LEXICILLHLTP 58

RESULT      2
ID PCT-US93-08528-51 STANDARD; PRT; 325 AA.
XX xxxxxx
AC AC
AA AC
DT DT
DX DX
DE DE
XX Sequence 51, Application PC/TUS9308528
CC Sequence 51, Application PC/TUS9308528
CC GENERAL INFORMATION:
CC APPLICANT: New York University
CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC NUMBER OF SEQUENCES: 348
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BROWDY AND NEWMARK
CC STREET: 419 Seventh Street, N.W., Suite 300
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20004
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/08528
CC FILING DATE: 09-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/943,236
CC FILING DATE: 10-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Townsend, Kevin G.
CC REGISTRATION NUMBER: 34,033
CC REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-628-5197
CC TELEFAX: 202-737-3528
CC TELEX: 248633
CC INFORMATION FOR SEQ ID NO: 51:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 325 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 325 AA; 3728 MW; 592694 CN;

Query Match      15.2%; Score 85; DB 3; Length 325;
Best Local Similarity 44.4%; Pred. No. 2.28e+01;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db    27 YMKLKYASYVFLNALADLCFLTLTP 53
      ::|:|||::|||::|::|::|::|::|
QY    34 FLTNSSYSESL-EL-LEXICILLHLTP 58
```

```

Sequence 8, Application PC/TUS9305704
Sequence 8, Application PC/TUS9305704
GENERAL INFORMATION:
APPLICANT: Edwards, Robert H
TITLE OF INVENTION: Vesicle Membrane Transport Proteins
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEES: Sheldon & Mak
STREET: 225 South Lake Avenue, Ninth Floor
CITY: Pasadena
STATE: California
COUNTRY: USA
ZIP: 91001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
. COMPUTER: IBM PC Compatible

```

CC	OPERATING SYSTEM:	PC-DOS/MS-DOS	
CC	SOFTWARE:	Patentin Release #1.0, Version #1.25	
CC	CURRENT APPLICATION DATA:		
CC	APPLICATION NUMBER:	PCT/US93/05704	
CC	FILING DATE:	19930611	
CC	CLASSIFICATION:		
CC	ATTORNEY/AGENT INFORMATION:		
CC	NAME:	Farber, Michael B	
CC	REGISTRATION NUMBER:	32,612	
CC	REFERENCE/DOCKET NUMBER:	9067-1PCT	
CC	TELECOMMUNICATION INFORMATION:		
CC	TELEPHONE:	(818) 796-4000	
CC	TELEFAX:	(818) 795-6321	
CC	INFORMATION FOR SEQ ID NO:	8:	
CC	SEQUENCE CHARACTERISTICS:		
CC	LENGTH:	194 amino acids	
CC	TYPE:	AMINO ACID	
CC	TOPOLOGY:	linear	
CC	MOLECULE TYPE:	peptide	
CC	HYPOTHETICAL:	NO	
CC	FRAGMENT TYPE:	internal	
CC	ORIGINAL SOURCE:		
CC	ORGANISM:	Transposon 10	
CC	SEQUENCE	194 AA; 20868 MW; 201442 CN;	
CC	Query Match	14.9%; Score 83; DB 3; Length 194;	
CC	Best Local Similarity	31.0%; Pred.No. 3.14e+01;	
CC	Matches	18; Conservative 13; Mismatches 23; Indels 4; Gaps 4;	
DB	1	MNSSTKIALVITLLDAM-GIGLIMFVLPILLRE-FIASEDIANHFVLLA-LYALMOV 55	
QY	1	MGSGPLVALLTLGSSHGTFGWT-LQLKLKSELTNNSYESSFLELLERKLCILLHL 57	
CC	RESULT	5	
CC	ID	US-08-741-406-2	STANDARD; PRT; 169 AA.
CC	XX	xxxxxx	
CC	XX		
CC	XX		
CC	XX		
CC	XX	Sequence 2, Application US/08741406	
CC	XX		
CC	XX	Sequence 2, Application US/08741406	
CC	XX	Patent No. 5721119	
CC	XX	GENERAL INFORMATION:	
CC	XX	APPLICANT: Scheffler, Immo E.	
CC	XX	TITLE OF INVENTION: Mammalian Artificial Chromosomes and	
CC	XX	TITLE OF INVENTION: Methods of Using Same	
CC	XX	NUMBER OF SEQUENCES: 16	
CC	XX	CORRESPONDENCE ADDRESS:	
CC	XX	ADDRESSEE: Campbell & Flores LLP	
CC	XX	STREET: 4370 La Jolla Village Drive, Suite 700	
CC	XX	CITY: San Diego	
CC	XX	STATE: California	
CC	XX	COUNTRY: United States	
CC	XX	ZIP: 92122	
CC	XX	COMPUTER READABLE FORM:	
CC	XX	MEDIUM TYPE: Floppy disk	
CC	XX	COMPUTER: IBM PC compatible	
CC	XX	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	XX	SOFTWARE: Patentin Release #1.0, Version #1.25	
CC	XX	CURRENT APPLICATION DATA:	
CC	XX	APPLICATION NUMBER: US/08/741,406	
CC	XX	FILING DATE:	
CC	XX	CLASSIFICATION: 514	
CC	XX	PRIOR APPLICATION DATA:	
CC	XX	APPLICATION NUMBER: US 08/550,717	
CC	XX	FILING DATE: 31-OCT-1995	
CC	XX	ATTORNEY/AGENT INFORMATION:	
CC	XX	NAME: Campbell, Cathryn A.	
CC	XX	REGISTRATION NUMBER: 31,815	
CC	XX	REFERENCE/DOCKET NUMBER: P-UD 2317	

Query Match	Best Local Similarity	Score	DB 1	Length	DB 2	Length	DB 3	Length	DB 4	Length	DB 5	Length	DB 6	Length	DB 7	Length	DB 8	Length	DB 9	Length	DB 10	Length	DB 11	Length	DB 12	Length	DB 13	Length	DB 14	Length	DB 15	Length	DB 16	Length	DB 17	Length	DB 18	Length	DB 19	Length	DB 20	Length	DB 21	Length	DB 22	Length	DB 23	Length	DB 24	Length	DB 25	Length	DB 26	Length	DB 27	Length	DB 28	Length	DB 29	Length	DB 30	Length	DB 31	Length	DB 32	Length	DB 33	Length	DB 34	Length	DB 35	Length	DB 36	Length	DB 37	Length	DB 38	Length	DB 39	Length	DB 40	Length	DB 41	Length	DB 42	Length	DB 43	Length	DB 44	Length	DB 45	Length	DB 46	Length	DB 47	Length	DB 48	Length	DB 49	Length	DB 50	Length	DB 51	Length	DB 52	Length	DB 53	Length	DB 54	Length	DB 55	Length	DB 56	Length	DB 57	Length	DB 58	Length	DB 59	Length	DB 60	Length	DB 61	Length	DB 62	Length	DB 63	Length	DB 64	Length	DB 65	Length	DB 66	Length	DB 67	Length	DB 68	Length	DB 69	Length	DB 70	Length	DB 71	Length	DB 72	Length	DB 73	Length	DB 74	Length	DB 75	Length	DB 76	Length	DB 77	Length	DB 78	Length	DB 79	Length	DB 80	Length	DB 81	Length	DB 82	Length	DB 83	Length	DB 84	Length	DB 85	Length	DB 86	Length	DB 87	Length	DB 88	Length	DB 89	Length	DB 90	Length	DB 91	Length	DB 92	Length	DB 93	Length	DB 94	Length	DB 95	Length	DB 96	Length	DB 97	Length	DB 98	Length	DB 99	Length	DB 100	Length	DB 101	Length	DB 102	Length	DB 103	Length	DB 104	Length	DB 105	Length	DB 106	Length	DB 107	Length	DB 108	Length	DB 109	Length	DB 110	Length	DB 111	Length	DB 112	Length	DB 113	Length	DB 114	Length	DB 115	Length	DB 116	Length	DB 117	Length	DB 118	Length	DB 119	Length	DB 120	Length	DB 121	Length	DB 122	Length	DB 123	Length	DB 124	Length	DB 125	Length	DB 126	Length	DB 127	Length	DB 128	Length	DB 129	Length	DB 130	Length	DB 131	Length	DB 132	Length	DB 133	Length	DB 134	Length	DB 135	Length	DB 136	Length	DB 137	Length	DB 138	Length	DB 139	Length	DB 140	Length	DB 141	Length	DB 142	Length	DB 143	Length	DB 144	Length	DB 145	Length	DB 146	Length	DB 147	Length	DB 148	Length	DB 149	Length	DB 150	Length	DB 151	Length	DB 152	Length	DB 153	Length	DB 154	Length	DB 155	Length	DB 156	Length	DB 157	Length	DB 158	Length	DB 159	Length	DB 160	Length	DB 161	Length	DB 162	Length	DB 163	Length	DB 164	Length	DB 165	Length	DB 166	Length	DB 167	Length	DB 168	Length	DB 169	Length	DB 170	Length	DB 171	Length	DB 172	Length	DB 173	Length	DB 174	Length	DB 175	Length	DB 176	Length	DB 177	Length	DB 178	Length	DB 179	Length	DB 180	Length	DB 181	Length	DB 182	Length	DB 183	Length	DB 184	Length	DB 185	Length	DB 186	Length	DB 187	Length	DB 188	Length	DB 189	Length	DB 190	Length	DB 191	Length	DB 192	Length	DB 193	Length	DB 194	Length	DB 195	Length	DB 196	Length	DB 197	Length	DB 198	Length	DB 199	Length	DB 200	Length	DB 201	Length	DB 202	Length	DB 203	Length	DB 204	Length	DB 205	Length	DB 206	Length	DB 207	Length	DB 208	Length	DB 209	Length	DB 210	Length	DB 211	Length	DB 212	Length	DB 213	Length	DB 214	Length	DB 215	Length	DB 216	Length	DB 217	Length	DB 218	Length	DB 219	Length	DB 220	Length	DB 221	Length	DB 222	Length	DB 223	Length	DB 224	Length	DB 225	Length	DB 226	Length	DB 227	Length	DB 228	Length	DB 229	Length	DB 230	Length	DB 231	Length	DB 232
-------------	-----------------------	-------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

[illegible]

```
CC ADDRESSEE: Cooper & Dunham
CC STREET: 30 Rockefeller Plaza
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect, Version 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/117,006
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 1795/39318
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-977-9550
CC TELEFAX: 212-664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 390 amino acids
CC TYPE: AMINO ACID
CC STRANDEDNESS: unknown
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC IMMEDIATE SOURCE:
CC CLONE: 5-HT1DB
CC SEQUENCE 390 AA; 43656 MW; 849817 CN;

Query Match 14.3%; Score 80; DB 1; Length 390;
Best Local Similarity 23.6%; Pred. No. 5.04e+01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;

Db 52 LVMLALITLTSNAFVATVTRKLTPTANYLIASLDVTDLLVSLVPISTMTV 111
QY 7 LVLLTLGGSHGTGCMTQLKLESFL-TNSSYESSFLELLEKICLLHLPSTSVTL 65
Db 112 TDRWTLQVVD 123
QY 66 HHARSQHRVVCN 77

RESULT 12
ID PCT-US93-00149-6 STANDARD; PRT; 390 AA.
AC xxxxxx
DI
DE
Sequence 6, Application PC/TUS9300149
Sequence 6, Application PC/TUS9300149
GENERAL INFORMATION:
APPLICANT: Weinshank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

CC ADDRESSEE: Cooper & Dunham
CC STREET: 30 Rockefeller Plaza
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect, Version 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/117,006
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 1795/39318
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-977-9550
CC TELEFAX: 212-664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 390 amino acids
CC TYPE: AMINO ACID
CC STRANDEDNESS: unknown
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC IMMEDIATE SOURCE:
CC CLONE: 5-HT1DB
CC SEQUENCE 390 AA; 43656 MW; 849817 CN;

Query Match 14.3%; Score 80; DB 1; Length 390;
Best Local Similarity 23.6%; Pred. No. 5.04e+01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;

Db 52 LVMLALITLTSNAFVATVTRKLTPTANYLIASLDVTDLLVSLVPISTMTV 111
QY 7 LVLLTLGGSHGTGCMTQLKLESFL-TNSSYESSFLELLEKICLLHLPSTSVTL 65
Db 112 TDRWTLQVVD 123
QY 66 HHARSQHRVVCN 77

RESULT 13
ID US-08-216-594-6 STANDARD; PRT; 390 AA.
AC xxxxxx
DI
DE
Sequence 6, Application US/08216594
Sequence 6, Application US/08216594
Patent No. 5652113
GENERAL INFORMATION:
APPLICANT: Weinshank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR
AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```



```

CC REFERENCE/DOCKET NUMBER: 1795/39317-22/JPW/NAT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-278-0400
CC TELEX: 212-391-0525
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 398 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
SQ SEQUENCE 398 AA; 44384 MW; 880684 CN;

Query Match 14.3%; Score 80; DB 1; Length 398;
Best Local Similarity 23.6%; Pred. No. 5,04e-01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;

Db 52 LVMLIALTLATLTSNAFVIATVYTRKRLTHANLYASLDYLLYSIIVIPISITMTTV 111
||:||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 7 LVLLTLTGSGHGTPGWTQLQKUESFL-TWSSVESSFELEKICLLHLPGSGTSTVL 65
||:||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 112 TDRWTLISQVCD 123
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 HHARSQHHVCN 77
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: Mon Mar 20 12:33:28 2000
Job time : 9 secs.

(TM)

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	%				
1	101	18.1	501	1	C70384	protein export membra	8.92e-02
2	93	16.7	382	2	S3638	ubiquinol - cytochrome	9.25e-01
3	92	16.5	712	2	A48156	translation regulator	1.23e+00
4	91	16.3	252	2	J95237	osmotin-like protein	1.63e+00
5	89	15.3	875	2	S73757	hypothetical protein	2.85e+00
6	89	15.3	3005	2	S36642	hemocytic protein zfh-	2.85e+00
7	88	15.8	314	2	B75076	daunorubicin resistan	3.76e+00
8	88	15.8	1822	2	S44849	K12H4.8 protein - Cae	3.76e+00
9	87	15.6	143	2	S43071	hypothetical protein	4.95e+00
10	87	15.6	396	2	S58161	probable translation	4.95e+00
11	86	15.4	272	2	S39641	flagellar motor appar	6.50e+00
12	85	15.2	315	2	H64082	hypothetical protein	8.53e+00
13	84	15.1	441	2	T74800	dihydroorotase (bc 3,	1.11e-01
14	84	15.1	454	2	R03130	probable tyrosine kin	1.11e-01
15	84	15.1	509	2	S76731	hypothetical protein	1.11e-01
16	83	14.9	378	2	D70324	hypothetical protein	1.45e+01
17	83	14.9	386	2	K14243	cytochrome b - dermat	1.45e+01
18	83	14.9	401	1	Y75C70	tetracycline resistan	1.45e+01
19	83	14.3	469	2	G65058	hypothetical protein	1.45e+01
20	83	14.9	647	2	A37086	beta-galactosidase (E	1.45e+01
21	82	14.7	56	2	S53000	mitotic-specific cycl	1.89e+01
22	82	14.7	110	2	S22896	T-cell receptor alpha	1.89e+01
23	82	14.7	110	2	S23368	T-cell receptor alpha	1.89e+01


```

##molecule_type DNA
##residues 1-879 ##label HIM
##cross-references EMBL:AE000042; GB:U00089; NID:g174112; PID:g174117
##note the nucleotide sequence was submitted to the EMBL Data
Library, November 1996

```

```

GENETICS
#genetic_code SGC3
SUMMARY
#length 879 #molecular-weight 101086 #checksum 8208

Query Match      15.9%; Score 89; DB 2; Length 879;
Best Local Similarity 28.9%; Pred. No. 2.85e+00;
Matches 11; Conservative 15; Mismatches 11; Indels 1; Gaps 1;

Db 677 EGIPKDSNY-SSEVHLDDQKSLFLQLARVSGDINENK 713
QY 32 ESFLNYSVESFELEELLCILLHLPSTGTSVTLHHR 69

```

RESULT	6
ENTRY	S33642
TITLE	homeotic protein zfh-2 - fruit fly (<i>Drosophila melanogaster</i>)
ORGANISM	#formal_name <i>Drosophila melanogaster</i>
DATE	20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1999
ACCESSIONS	S33642; S27817
REFERENCE	S33641
#authors	Fortini, M.E.; Lai, Z.; Rubin, G.M.
#journal	Mech. Dev. (1991) 34:113-122
#title	The <i>Drosophila</i> zfh-1 and zfh-2 genes encode novel proteins containing both zinc-finger and homeodomain motifs.

cross-references MUID:92001539
 accession S33642 preliminary
 ##status
 ##molecule_type mRNA
 ##residues 1-3005 ##label FOR
 ##cross-references EMBL:M63450; NID:g15882; PID:AAA29051.1;
 PMID:1508323

GENETICS
#gene
zfh-2
##cross-references FlyBase:FBp0004607
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation;
zinc finger

```

FEATURE
1798-1854      #domain homeobox homology #label HOX1\
2155-2211      #domain homeobox homology #label HOX2\
2761-2817      #domain homeobox homology #label HOX3
#length 3005  #molecular-weight 332056 #checksum 7516
SUMMARY

```

Query Match	15.9%	Score 89;	DB 2;	Length 3005;
Best Local Similarity	38.0%	Pred. No. 2.85e+00;		
Matches	19;	Conservative	11;	Mismatches 17;
			Indels	3;
			Gaps	3;

Db 2165 QIKVLQEFFENNYPKSDLELYLSKL-LLLS-PRVIVVFQNARQORKI 2212
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 27 QLKIKESFLTNSSY-ESSFLELLEKICLLHLHPSGTSWTLHHARSQHVV 75

RESULT	7
ENTRY	B75076
TITLE	daunorubicin resistance ATP-binding protein drra PAB1703 - Pyrococcus abyssi (strain Orsay)
ORGANISM	Pyrococcus abyssi
DATE	20-Aug-1999
ACCESSIONS	B75076
REFERENCE	A75001
authors	anonymous, Genoscope
submission	submitted to the EMBL Data Library, July 1999
description	Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.
#accession	B75076
##status	preliminary

```

#molecule_type DNA
#residues 1-314 #label RAW
#cross-references GB:AJ248286; GB:AL096836; NID:G5458356;
PIDN:ICAB49911.1; PID:e151808; PID:g5458423
#experimental_source strain Orsay

```

```

GENETICS
#gene
SUMMARY      PAB1703
               #length 314 #molecular-weight 35594 #checksum 3877

Query Match      15.88; Score 88; DB 2: Length 314;
Best Local Similarity 33.38; Pred. No. 3.76e+00;
Matches      8; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

Db      291 ELKIKDIFTANFEDVYLEITER 314
QY      27 QLKIKSFTLNSVSSFFLEK 50

```

RESULT	8
ENTRY	
TITLE	S44849 #type complete
ORGANISM	K12H4.8 protein - Caenorhabditis elegans
DATE	#formal_name Caenorhabditis elegans 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 20-Sep-1999
ACCESSIONS	S44849
REFERENCE	S44617
#authors	Favell, A.D.
#submission	submitted to the EMBL Data Library, May 1993
#description	Sequence of the c. elegans cosmid K12H4.
#accession	S44849

```
##status Preliminary
##molecule_type DNA
##residues 1-1822 #label FAV
##cross-references EMBL:L14331; NID:q289702; PDB:AAA38101.1;
PID:q289703
```

GENETICS
#Introns
17/3; 72/1; 116/3; 170/1; 216/3; 340/3; 401/3; 444/3; 512/2;
634/3; 694/3; 745/1; 855/2; 1053/1; 1113/3; 1215/3; 1321/1;
1396/2; 1458/3; 1507/3; 1563/2; 1645/3; 1677/3; 1804/2
#superfamily unassigned DEAD/H box helicases; DEAD/H box

CLASSIFICATION

```

KEYWORDS
FEATURE
33-496 #domain DEAD/H box helicase homology #label DEAD\
33-40 #region nucleotide-binding motif A (P-loop)\
141-146 #region nucleotide-binding motif B\
145-148 #region DEXH motif
#length 1822 #molecular-weight 208291 #checksum 7039

SUMMARY

```

Query Match 15.8%; Score 88; DB 2; Length 1822;
Best Local Similarity 31.4%; Pred. No. 3.76e+00;
Matches 16: Conservative 17; Mismatches 15; Indels 3; Gaps 3;

DB 1300 IGLGVSPCLLITALTTSNAAD-GMSLERFETIGDSFLKFATTDVLYHTLLD 1349
: : : : | : : : | : : : : : : : : : :
QY 1 MGSGPLVLTLTLLGSSHGTCPGMTLO-L-KLKESILTNSSESELELLE 49

RESULT	9
ENTRY	S43071
TITLE	#type complete
ORGANISM	hypothetical protein 5 - human herpesvirus 6
DATE	#formal_name human herpesvirus 6
	19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change
	17-Mar-1999
ACCESSIONS	S43071
REFERENCE	S43067
#authors	Thompson, J.; Choudhury, S.; Kashanchi, F.; Doniger, J.; Berneman, Z.; Frenkel, N.; Rosenthal, L.J.
#journal	Oncogene (1994) 9:1167-1175
#title	A transforming fragment within the direct repeat region of human herpesvirus type 6 that transactivates HIV-1.
#cross-references	NID:94181269
#accession	S43071

```

#status      nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues    1-143 #label THO
#cross-references EMBL:X73675; NID:g469952; PID:g469957
#note        the nucleotide sequence was submitted to the EMBL Data
              Library, June 1993
SUMMARY      #length 143 #molecular-weight 13317 #checksum 6533
Query Match  15.68; Score 87; DB 2; Length 143;
Best Local Similarity 47.28; Pred. No. 4.95e+00;
Matches 17; Conservative 4; Mismatches 14; Indels 1; Gaps 1;
Db 88 LGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 123
   :|||:|||||:|||||:|||||:|||||:
QY 1 MGSGLPLVLLTL-LGSSHGTPGCMTLQLKESFL 35

RESULT 10
ENTRY   S58161 #type complete
TITLE   Probable translation releasing factor RF-1 - fission yeast
        (Schizosaccharomyces pombe)
ALTERNATE_NAMES SPAC2F7.17 protein
ORGANISM  #formal_name Schizosaccharomyces pombe
DATE      13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
        26-Aug-1999
ACCESSIONS S58161
REFERENCE  S58145
#authors   Gentles, S.; Churcher, C.M.
#submission submitted to the EMBL Data Library, July 1995
#accession S58161
#status    preliminary
#molecule_type DNA
#residues  1-396 #label GEN
#cross-references EMBL:Z50142; NID:g1052783; PIDN:CAA90504.1;
              PID:g1052800
GENETICS
#introns   57/3; 90/3
CLASSIFICATION #superfamily translation releasing factor
SUMMARY      #length 396 #molecular-weight 44954 #checksum 9362
Query Match  15.68; Score 87; DB 2; Length 396;
Best Local Similarity 45.08; Pred. No. 4.95e+00;
Matches 9; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
Db 281 LTHIPTGTVSNQSRSQHQ 300
   |||:|||||:|||||:|||||:
QY 54 LLHLPSTSVTLHARSQHH 73

RESULT 11
ENTRY   S39641 #type complete
TITLE   flagellar motor apparatus homolog ytxD - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE      08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change
        23-Sep-1999
ACCESSIONS S39641; B70003
REFERENCE  S39641
#authors   Grundy, F.J.; Waters, D.A.; Takova, T.Y.; Henkin, T.M.
#journal   Mol. Microbiol. (1993) 10:259-271
#title     Identification of genes involved in utilization of acetate
              and acetoin in Bacillus subtilis.
#cross-references MUID:95020526
#accession S39641
#status    preliminary
#molecule_type DNA
#residues  1-272 #label GRU
REFERENCE  A69580
#authors   Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
              Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
              Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
              A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
              Bruschl, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
              Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;

```

```

Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.I.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lezarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Melliado, R.P.; Mizuno,
M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Taconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal   Nature (1997) 390:249-256
#title     The complete genome sequence of the Gram-positive bacterium
              Bacillus subtilis.
#cross-references MUID:98044033
#accession B70003
#status    preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues  1-272 #label KUN
#cross-references GB:299119; GB:AL009126; NID:g2635411;
              PIDN:CAB14951.1; PID:ell85846; PID:g4635457
#experimental_source strain 168
GENETICS
#gene      ytxD
CLASSIFICATION #superfamily flagellar motor rotation protein
              transmembrane protein
KEYWORDS    #length 272 #molecular-weight 30143 #checksum 4312
SUMMARY
Query Match  15.48; Score 86; DB 2; Length 272;
Best Local Similarity 34.68; Pred. No. 6.50e+00;
Matches 18; Conservative 12; Mismatches 20; Indels 2; Gaps 1;
Db 164 IGTVLGLVLMKLNLDPHMGNMAYALLTLYGSLANVMFNP1AKLEEK 215
   :|||:|||||:|||||:|||||:|||||:
QY 1 MGSGLPLVLLTLGSSHGTPGNTQL--KLKESFLNSSVSEFFLEK 50

RESULT 12
ENTRY   H64082 #type complete
TITLE   hypothetical protein HU630 - Haemophilus influenzae (strain
              Rd KW20)
ORGANISM  #formal_name Haemophilus influenzae
DATE      18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
        24-Sep-1999
ACCESSIONS H64082
REFERENCE  A64000
#authors   Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
              Kirkness, E.F.; Krievage, A.R.; Bult, C.J.; Tomb, J.F.;
              Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
              FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
              Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
              J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
              M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
              D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,

```


Search completed: Mon Mar 20 12:32:35 2000
Job time : 14 secs.

M P S R C H _ P P
***** (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 20 12:31:17 2000; MasPar time 5.65 Seconds
Tabular output not generated. 412.201 Million cell updates/sec

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect score: 538
Sequence: 1 MGSGLPLVLLTLGSSHGT.....SQTSTLHARSHQHVYCVNT 78

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 40.128; Variance 78.060; scale 0.514

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	101	18.1	501	1	SECD_AQUAE PROTEIN-EXPORT MEMBRAN	2.16e+02
2	92	16.5	712	1	EBZE_YEAST TRANSLATION INITIATION	3.73e-01
3	91	16.3	252	1	OLP1_LYCES OSMOTIN-LIKE PROTEIN P	5.07e-01
4	89	15.9	3005	1	ZFH2_DROME ZINC-FINGER PROTEIN 2	9.30e-01
5	88	15.8	1822	1	YH68_CABEL HYPOTHETICAL HELICASE	1.26e+00
6	87	15.6	396	1	RFIM_SCHPO PUTATIVE MITOCHONDRIAL	1.69e+00
7	86	15.4	272	1	YTXD_BAGSU HYPOTHETICAL 30.1 KD P	2.27e+00
8	85	15.2	315	1	RSEB_HAEIN SIGMA-E FACTOR REGULAT	3.04e+00
9	84	15.1	191	1	YDB3_SCHPO HYPOTHETICAL 21.1 KD P	4.07e+00
10	83	14.9	401	1	TCR2_ECOLI TETRACYCLINE RESISTANC	5.43e+00
11	83	14.9	445	1	YGC5_ECOLI HYPOTHETICAL METABOLIT	5.43e+00
12	83	14.9	647	1	BGAL_MOUSE BETA-GALACTOSIDASE PRE	5.43e+00
13	83	14.9	1233	1	NEE3_HUMAN GLUTAMATE [NMDA] RECEP	5.43e+00
14	82	14.7	169	1	C560_HUMAN SUCCINATE DEHYDROGENAS	7.22e+00
15	82	14.7	372	1	INVE_SALTY INVASION PROTEIN INVE.	7.22e+00
16	82	14.7	385	1	CYB_ASPNG CYTOCHROME B.	7.22e+00
17	81	14.5	253	1	MOTA_RHOSH CHEMOTAXIS MOTA PROTEI	9.58e+00
18	81	14.5	261	1	ASPK_MOUSE ACROSOMAL PROTEIN SP-1	9.58e+00
19	81	14.5	346	1	GP41_HUMAN PUTATIVE G PROTEIN-COU	9.58e+00
20	81	14.5	346	1	GP42_HUMAN PUTATIVE G PROTEIN-COU	9.58e+00
21	81	14.5	386	1	CYB_SARGL CYTOCHROME B.	9.58e+00
22	81	14.5	387	1	CYB_PODAN CYTOCHROME B.	9.58e+00
23	81	14.5	482	1	LBP_RABIT LIPOPOLYSACCHARIDE-BIN	9.58e+00

24	81	14.5	5255	1	BACA_BACLI BACITRACIN SYNTHETASE	9.58e+00
25	80	14.3	169	1	C560_BOVIN SUCCINATE DEHYDROGENAS	1.27e+01
26	80	14.3	191	1	Y064_TREPA HYPOTHETICAL PROTEIN T	1.27e+01
27	80	14.3	387	1	CYB_EMENI CYTOCHROME B.	1.27e+01
28	80	14.3	400	1	TCR8_PASMU TETRACYCLINE RESISTANC	1.27e+01
29	80	14.3	402	1	OPDE_PSEAE TRANSCRIPTION REGULATO	1.27e+01
30	80	14.3	411	1	VGLM_HSVBC GLYCOPROTEIN M.	1.27e+01
31	80	14.3	488	1	MB11_EMENI MENA NUTRIASE B11 (COB	1.27e+01
32	80	14.3	713	1	CDGT_BACSP CYCLOMALTOGENIN GLUC	1.27e+01
33	80	14.3	772	1	TF11_HUMAN NFE2-RELATED FACTOR 1	1.27e+01
34	79	14.2	259	1	ATP6_YEAST ATP SYNTHASE A CHAIN P	1.67e+01
35	79	14.2	393	1	CYB_VENIN CYTOCHROME B.	1.67e+01
36	79	14.2	408	1	BTN1_YEAST BTN1 PROTEIN.	1.67e+01
37	79	14.2	413	1	RFIM_YEAST MITOCHONDRIAL PEPTIDE	1.67e+01
38	79	14.2	493	1	ACHE_MOUSE ACETYLCHOLINE RECEPTOR	1.67e+01
39	79	14.2	505	1	ACHE_BOVIN ACETYLCHOLINE RECEPTOR	1.67e+01
40	79	14.2	511	1	MYIN_ECOLI VIRULENCE FACTOR MWIN	1.67e+01
41	79	14.2	562	1	GARP_HUMAN GARP PROTEIN PRECURSOR	1.67e+01
42	79	14.2	1015	1	PTPX_HUMAN PROTEIN-TYROSINE PHOSP	1.67e+01
43	78	14.0	101	1	GRO_CRIGR GROWTH REGULATED PROTE	2.20e+01
44	78	14.0	267	1	YTXD_BACNE HYPOTHETICAL 29.3 KD P	2.20e+01
45	78	14.0	1239	1	NME3_MOUSE GLUTAMATE [NMDA] RECEP	2.20e+01

ALIGNMENTS

RESULT 1
ID SECD_AQUAE STANDARD; PRT; 501 AA.
AC O67102;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PROTEIN-EXPORT MEMBRANE PROTEIN SECD.
GN SECD OR AQ_973.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GRATERLAND T., YOUNG W.G., LENOX A.L., GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R., RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus";
RL Nature 392:353-358(1998).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF000716; AAC07060.1; -
KW Protein transport; Translocation; Transmembrane; Membrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 339 359 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT TRANSMEM 470 490 POTENTIAL.
SQ SEQUENCE 501 AA; 55459 MW; E67F690C CRC32;
Query Match 18.1%; score 101; DB 1; Length 501;
Best Local Similarity 38.1%; Pred. No. 2.16e-02;
Matches 18; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

Db 454 VILFPG-CGSPVKGFTATLGTIASPISNYYAKVFLDLNLSKIL 499
 QY 8 VILLTLLGSSHGTPGNTQLKLESLTSSYESSFLELLEKLCIL 54

RESULT 2
 ID E2BE YEAST STANDARD; PRT; 712 AA.
 AC P32501;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DE TRANSLATION INITIATION FACTOR EIF-2B EPSILON SUBUNIT (EIF-2B GDP-GTP
 EXCHANGE FACTOR) (GUANINE NUCLEOTIDE EXCHANGE FACTOR SUBUNIT GCD6)
 DE (GCD COMPLEX SUBUNIT GCD6).
 GN GCD6 OR TIF225 OR YDR211W OR YD8142.12 OR YD8142B.03.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93180841.
 RA BUSHMAN J.L., ASURU A.I., MATTS R.L., HINNEBUSCH A.G.;
 RT "Evidence that GCD6 and GCD7, translational regulators of GCN4, are
 subunits of the guanine nucleotide exchange factor for eif-2 in
 Saccharomyces cerevisiae";
 RL Mol. Cell. Biol. 13:1920-1932(1993).
 RP SEQUENCE FROM N.A.
 RC STRAIN=S488C / AB972;
 RA OLIVER K., SHORE L., HARRIS D., BARRELL B.G., RAJANDREAM M.A.,
 RA WALSH S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SUBUNIT OF THE GUANINE NUCLEOTIDE EXCHANGE FACTOR FOR
 EIF-2. REQUIRED TO REPRESS GCN4 TRANSLATION UNDER NONSTARVATION
 CONDITIONS. GCD6 AND GCD7 REPRESS GCN4 EXPRESSION AT THE
 TRANSLATIONAL LEVEL BY ENSURING THAT RIBOSOMES WHICH HAVE
 TRANSLATED UORF1 WILL REINITIATE AT UORF2, -3, OR -4 AND THUS FAIL
 TO REACH THE GCN4 START SITE.
 CC -1- SUBUNIT: COMPLEX OF FIVE DIFFERENT SUBUNITS; ALPHA (GCN3), BETA
 (GCD7), GAMMA (GCD1), DELTA (GCD2) AND EPSILON (GCD6).
 CC -1- SIMILARITY: BELONGS TO THE EIF-2B GAMMA/EPSILON SUBUNITS FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L07115; AA65498.1; -;
 DR EMBL; Z68194; CA922354.1; -;
 DR EMBL; Z68195; CA922362.1; -;
 DR PIR; S30776; S30776.
 DR PIR; A48156; A48156.
 DR SGD; L0000674; GCD6.
 DR PFAM; PF00132; hexapep; 3.
 KW Amino-acid biosynthesis; Translation regulation.
 SQ SEQUENCE 712 AA; 81161 MW; 5DAD189F CRC32;

Query Match 16.5%; Score 92; DB 1; Length 712;
 Best Local Similarity 39.5%; Pred. NO. 3.73e-01;
 Matches 17; Conservative 8; Mismatches 14; Indels 4; Gaps 4;

Db 12 LGN-HGKSDMDVEDRQAQVLLDS-YETREMPITAVKPRCLL 52
 QY 14 LGSSHGTPGNTQLKLESLTSSYESSFLELLE-KL-CIL 54

RESULT 3
 ID OLPL_LYCES STANDARD; PRT; 252 AA.
 AC Q41350;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE OSMOTIN-LIKE PROTEIN PRECURSOR.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. VF36;
 RX MEDLINE; 97128324.
 RA CHEN R., WANG F., SMITH A.G.;
 RT "A flower-specific gene encoding an osmotin-like protein from
 Lycopersicon esculentum";
 RL Gene 179:301-302(1996).
 RN [2]
 RP SEQUENCE OF 25-33.
 RA ROBERTSON D., MITCHELL G.P., GILROY J.S., GERRISH C., BOLWELL G.P.,
 RA SLABAS A.R.;
 RL Submitted (JAN-1997) to the SWISS-PROT data bank.
 CC -1- SUBCELLULAR LOCATION: CELL WALL PROTEIN (POTENTIAL).
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L76632; AB41124.1; -;
 DR HSP; P25871; IAUH.
 DR PROSITE; PS00316; THAUMATIN; 1.
 DR PFAM; PF00314; thaumatin; 1.
 KW Cell wall; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 252 OSMOTIN-LIKE PROTEIN.
 SQ SEQUENCE 252 AA; 27265 MW; 229E5542 CRC32;

Query Match 16.3%; Score 91; DB 1; Length 252;
 Best Local Similarity 54.2%; Pred. NO. 5.07e-01;
 Matches 13; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Db 10 LPLSLFTLLSLSQSTNPFIILT 33
 QY 5 LPLVLLTLLGSSHGTPGNTQL 28
 ||| ||||| :||| | | |
 ||| ||||| :||| | | |

RESULT 4
 ID ZFH2_DROME STANDARD; PRT; 3005 AA.
 AC P28167;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DE ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
 GN ZFH-2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92001539.
 RA FORTINI M.E., LAI Z., RUBIN G.M.;
 RT "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
 containing both zinc-finger and homeodomain motifs";
 RL Mech. Dev. 34:113-122(1991).
 CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
 CC NERVOUS SYSTEM.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- ISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.

CC -!- SIMILARITY: CONTAINS THREE HOMEODX DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M63450; AA29051.1; --
 CC PIR: S27817; S27817
 CC PIR: S33642; S33642
 CC HSP: P15822; 4ZNF.
 CC TRANSFAC: T00920; --
 CC FLYBASE: FBgn004607; zfh2.
 CC PROSITE: PS00027; HOMEODX_1; 2.
 CC PROSITE: PS00028; ZINC_FINGER_C2H2; 8.
 CC PROSITE: PS00071; HOMEODX_2; 3.
 CC PFAM: PF00046; homeobox; 3.
 CC PFAM: PF00096; zf-C2H2; 12.
 CC Zinc-finger; Metal-binding; DNA-binding; Homeobox; Nuclear protein;
 KW Repeat:
 KW ZN_FING 133 156 C2H2-TYPE.
 KW ZN_FING 559 582 C2H2-TYPE.
 KW ZN_FING 614 638 C2H2-TYPE.
 KW ZN_FING 732 756 C2H2-TYPE.
 KW ZN_FING 897 916 C2H2-TYPE (DEGENERATE).
 KW ZN_FING 940 964 C2H2-TYPE.
 KW ZN_FING 1023 1039 C2H2-TYPE.
 KW ZN_FING 1074 1098 C2H2-TYPE.
 KW ZN_FING 1210 1233 C2H2-TYPE.
 KW ZN_FING 1341 1365 C2H2-TYPE.
 KW ZN_FING 1438 1462 C2H2-TYPE.
 KW ZN_FING 1477 1500 C2H2-TYPE (DEGENERATE).
 KW ZN_FING 1513 1535 C2H2-TYPE.
 KW ZN_FING 1541 1564 C2H2-TYPE.
 KW DNA_BIND 1797 1856 HOMEODX 1.
 KW DNA_BIND 2154 2213 HOMEODX 2.
 KW ZN_FING 2234 2256 C2H2-TYPE.
 KW ZN_FING 2371 2393 C2H2-TYPE.
 KW DNA_BIND 2760 2819 HOMEODX 3.
 KW SEQUENCE 3005 AA; 332056 MW; 8B4CC45F CRC32;
 SQ
 Query Match 15.9%; Score 89; DB 1; Length 3005;
 Best Local Similarity 38.0%; Pred. No. 9.30e-01;
 Matches 19; Conservative 11; Mismatches 17; Indels 3; Gaps 3;
 DB 2165 QIKVLQEFFENSYPKDSLEVLKLLLS-PRVIVWFQNAKQKQKI 2212
 QY 27 QLKKEFLTNSSY-ESSFLELLEKLLHLHPSGTSVTLHARSQHV 75
 RESULT 5
 ID YW68 CAEEL STANDARD; PRT; 1822 AA.
 AC P34529;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL HELICASE K12H4.8 IN CHROMOSOME III.
 GN K12H4.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoidea; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTDL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSIER N.,

RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHDWKKEEN R.,
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SDNNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MTEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATERSDN R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WHELDON P.;
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.;
 RL Nature 368:32-38(1994).
 CC -!- SIMILARITY: WITH OTHER ATP DEPENDENT HELICASES.
 CC -!- SIMILARITY: CONTAINS A RNASE III DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L14331; AAA28101.1; --
 CC PIR: S44849; S44849
 CC WORMPEP: K12H4.8; CE00273.
 CC PROSITE: PS00517; RIBONUCLEASE_III; 1.
 CC PFAM: PF00035; dsrm; 1.
 CC PFAM: PF00270; DEAD; 1.
 CC PFAM: PF00271; helicase_C; 1.
 CC PFAM: PF00636; Ribonuclease_3; 2.
 KW Hypothetical protein; Helicase; ATP-binding; Hydrolase; Nuclease;
 KW Endonuclease.
 KW NP_BIND 33 40 ATP (POTENTIAL).
 KW SITE 145 148 DECH BOX.
 FT DOMAIN 1534 1822 RNASE III DOMAIN.
 FT SEQUENCE 1822 AA; 208291 MW; 4F8E56BE CRC32;
 SQ
 Query Match 15.8%; Score 88; DB 1; Length 1822;
 Best Local Similarity 31.4%; Pred. No. 1.26e+00;
 Matches 16; Conservative 17; Mismatches 15; Indels 3; Gaps 3;
 DB 1300 IGLGVSPCLLLTALTSNAD-GMSLEFETIGDSFLKFKATDYLYHTLLD 1349
 QY 1 MGSGLPLVLLLLGSSHGPGWTQ-L-KLKEFLTNSSYESSFELLE 49
 RESULT 6
 ID RFLM SCHPO STANDARD; PRT; 396 AA.
 AC Q09691;
 DT 01-NDV-1995 (Rel. 32, Created)
 DT 01-NDV-1995 (Rel. 32, Last sequence update)
 DT 01-NDV-1995 (Rel. 32, Last annotation update)
 DE PUTATIVE MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR PRECURSOR.
 GN SPAC2F7.17.
 DS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA GENTLES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE PRKARYOTIC AND MITOCHONDRIAL RELEASE
 CC FACTORS FAMILY. STRONG TO YEAST MRF-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z50142; CAA90504.1; --

```

Db 164 ICTGVGLVLMKLNIDPHMLGPNMAIALLTLYGSLIANVFNPTAAKLEEK 215
      ::: |||:: | | | | | | | | | | | | | | | | | | | |
QY 1 MGSGLPVLVLLTLLGSGSHGTGPGMTQL--KLKESFLTNSVSESSFLELLEK 50

RESULT 8
ID RSEB HAEIN STANDARD; PRT; 315 AA.
AC P44792;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE SIGMA-E FACTOR REGULATORY PROTEIN RSEB HOMOLOG PRECURSOR.
DE RSEB OR HI0630.
GN Haemophilus influenzae.
OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-RD / KW20;
RC MEDLING: 95350630.
RX KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICKS
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODE A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.J., GREGG HAGEN N.S.M.,
RA GHEM C.L., McDONALD L.A., SMALL K.J., GREGG HAGEN N.S.M.,
RA VENTER J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.;
RL Science 269:496-512(1995).
CC -!- FUNCTION: SEEMS TO MODULATE THE ACTIVITY OF RPOE (SIGMA-E).
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
CC -!- SIMILARITY: TO E.COLI RSEB AND P.AERUGINOSA MUCB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a c
CC between the Swiss Institute of Bioinformatics and the EMBL
CC the European Bioinformatics Institute. There are no restricti
CC use by non-profit institutions as long as its content is
CC modified and this statement is not removed. Usage by and fo
CC entities requires a license agreement (See http://www.isb-sib
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: U32745; AAC22290.1; -
DR TIGR: HI0630; -
KW Periplasmic; Signal.
FT SIGNAL 1 23
FT CHAIN 24 315
FT FT
SQ SEQUENCE 315 AA; 35906 MW; 80B1C1ACD CRC32;
Query Match 15.2%; Score 85; DB 1; Length 315;
Best Local Similarity 29.5%; Pred. No. 3,04e+00;
Matches 13; Conservative 18; Mismatches 11; Indels 2;

Db 9 TALSLLSLSSIASAEELSAKQSLD-KWTQA-IDNINVEIFAPVQ 50
      ::: |||:: | | | | | | | | | | | | | | | | | | | |
QY 3 SGLPVLVLLTLLGSGSHGTGPGMTQLKLKESFLTNSVSESSFLE 46

RESULT 9
ID YDB3_SCHPO STANDARD; PRT; 191 AA.
AC Q10356;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 21.1 KD PROTEIN C32E12.03C IN CHROMOSOME 1.
DE SPAC22E12.03C.
GN Schizosaccharomyces pombe (Fission yeast).
OS Schizosaccharomycetes; Ascomycota; Archaescomycetes;
OC Eukaryota; Fungi; Ascomycota; Archaescomycetes;

```



```

Biochem. Biophys. Res. Commun. 178:158-164(1991)].
CC -|- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM
CC -| GLANGLOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCANs.
CC -|- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC -| GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -|- SUBCELLULAR LOCATION: LYSOSOMAL.
CC -|- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collabora-
tion between the Swiss Institute of Bioinformatics and the EMBL outsta-
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in n-
modified and this statement is not removed. Usage by and for commu-
nities requires a license agreement (See http://www.isb-sib.ch/ann
or send an email to licensel@isb-sib.ch).
-----
EMBL; M57734; AAA37293.1; .
DR EMBL; M75122; AAA37292.1; .
DR EMBL; M75107; AAA37292.1; JOINED
DR EMBL; M75108; AAA37292.1; JOINED.
DR EMBL; M75109; AAA37292.1; JOINED.
DR EMBL; M75111; AAA37292.1; JOINED.
DR EMBL; M75112; AAA37292.1; JOINED.
DR EMBL; M75113; AAA37292.1; JOINED.
DR EMBL; M75114; AAA37292.1; JOINED.
DR EMBL; M75115; AAA37292.1; JOINED.
DR EMBL; M75116; AAA37292.1; JOINED.
DR EMBL; M75117; AAA37292.1; JOINED.
DR EMBL; M75118; AAA37292.1; JOINED.
DR EMBL; M75119; AAA37292.1; JOINED.
DR EMBL; M75120; AAA37292.1; JOINED.
DR EMBL; M75121; AAA37292.1; JOINED.
DR FIR; A37086; A37086.
DR MGD; MG188151; BGL.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
DR PFAM; PF01301; Glyco_hydro_35; 1.
KW Hydrolase; Glycosidase; Lysosome; Signal; Glycoprotein.
FT SIGNAL 1 24
FT PROPEP 25 29
FT CHAIN 30 647
FT ACT_SITE 189 189
FT ACT_SITE 269 269
FT CARBOHYD 27 27
FT CARBOHYD 248 248
FT CARBOHYD 500 500
FT CARBOHYD 504 504
FT CARBOHYD 510 510
FT CARBOHYD 544 544
FT CARBOHYD 557 557
FT CARBOHYD 617 617
FT CONFLICT 517 517 N -> D (IN REF. 2).
FT CONFLICT 539 539 G -> R (IN REF. 2).
FT SEQUENCE 647 AA; 73121 MW; 15BCF158 CRC32;
Query Match 14.9%; Score 83; DB 1; Length 647;
Best Local Similarity 37.8%; Pred.No. 5.43e+00;
Matches 14; Conservative 8; Mismatches 12; Indels 3; Gaps
Db 10 LPFLALQLGAAHGIIYNVTORTFKLDYSDRFLKDG 46
|||: || ||::||| : | : | : | : | :
QY 5 LPVLVLLTLGGSSGT-G-PGWITLQLKL-KESFLTNS 38
-----
RESULT 13
ID NM33_HUMAN STANDARD; PRT; 1233 AA.
AC Q14957;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 3 PRECURSOR (N-METHYL
DE D-ASPARTATE RECEPTOR SUBTYPE 2C) (NR2C) (NMNDR2C).
GN GRIN2C.
```


OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN;
RX MEDLINE; 97189248.
RA LIN Y.J., BOVETTO S., CARVER J.M., GIORDANO T.;
RT "Cloning of the cDNA for the human NMDA receptor NR2C subunit and its
RT expression in the central nervous system and periphery.";
RL Brain Res. Mol. Brain Res. 43:57-64(1996).
CC -!- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC -!- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MAINLY IN BRAIN WITH PREDOMINANT EXPRESSION IS
CC IN THE CEREBELLUM, ALSO PRESENT IN THE HIPPOCAMPUS, AMYGDALA,
CC CAUDATE NUCLEUS, CORPUS CALLOSUM, SUBTHALAMIC NUCLEI AND THALAMUS.
CC DETECTED IN THE HEART, SKELETAL MUSCLE AND PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L76224; AAA8096.1; -.
CC MIM; 138254; -.
CC PFAM; PF00060; lig_chan; 1.
CC Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
CC Ionic channel; Magnesium.
CC SIGNAL 1 19
CC CHAIN 20 1233
CC -----
CC TRANSMEM 554 574
CC TRANSMEM 597 617
CC TRANSMEM 627 647
CC TRANSMEM 815 835
CC SITE 612 612
CC -----
CC CARBOHYD 70 70
CC CARBOHYD 337 337
CC CARBOHYD 438 438
CC CARBOHYD 539 539
CC SEQUENCE 1233 AA; 134239 MW; 671F9981 CRC32;
CC -----
CC Query Match 14.9%; Score 83; DB 1; Length 1233;
CC Best Local Similarity 33.3%; Pred. No. 5.43e+00;
CC Matches 20; Conservative 13; Mismatches 24; Indels 3; Gaps 3;
CC -----
CC Db 1 MCGALGPALLTLFGAGLGGGOGGGMVAVFSSSGPPQAOFRARLTPOSEFL-DLP 59
CC Qy 1 MGSGLPLVLLT-LGSSHGTEGGMTLQ-LKIKESFLTNSSVESSEFLEKLCILLHLP 58
CC -----
CC RESULT 14
CC ID C560_HUMAN STANDARD; PRT; 169 AA.
CC AC Q99643;
CC DT 15-JUL-1999 (Rel. 38, Created)
CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE SUCCINATE DEHYDROGENASE CYTOCHROME B560 SUBUNIT PRECURSOR (QPS1)
CC DE (CII-3) (SUCCINATE DEHYDROGENASE COMPLEX SUBUNIT C).
CC GN SDHC OR CYB560 OR SDH3.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
CC [1]
CC SEQUENCE FROM N.A.
CC -----
CC RA AU H.C., RAVAL P.J., SCHEFFLER I.E.;
CC "The cDNA sequence of human CII-3, an integral membrane protein
CC subunit of complex II of the mitochondria electron transport
CC chain.";
CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 98194224.
CC RA HIRAWAKE H., TANIWAKI M., TAMURA A., KOJIMA S., KITA K.;
CC "Cytochrome b in human complex II (succinate-ubiquinone
CC oxidoreductase): cDNA cloning of the components in liver mitochondria
CC and chromosomal assignment of the genes for the large (SDHC) and small
CC (SDHD) subunits to 1q21 and 11q23.";
CC RT Cytogenet. Cell Genet. 79:132-138(1997).
CC RL [3]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 98372071.
CC RA ELBEHI-GREEN A., AU H.C., MASCARELLO J.T., REAM-ROBINSON D.,
CC SCHEFFLER I.E.;
CC "Characterization of the human SDHC gene encoding one of the integral
CC membrane proteins of succinate-quinone oxidoreductase in
CC mitochondria.";
CC RL Gene 213:133-140(1998).
CC -!- FUNCTION: MONO-HEME CYTOCHROME B. MAY ACT AS A MEDIATOR OF LOW
CC POTENTIAL COUPLES IN AN ELECTRON FLOW THROUGH CARDIAC COMPLEX II.
CC IS INVOLVED IN SYSTEM II OF THE MITOCHONDRIAL ELECTRON TRANSPORT
CC CHAIN WHICH IS RESPONSIBLE FOR TRANSFERRING ELECTRONS FROM
CC SUCCINATE TO UBIQUINONE (COENZYME Q).
CC -!- SUBUNIT: FORMS PART OF COMPLEX II CONTAINING FOUR SUBUNITS: A 70
CC KD FLAVOPROTEIN (FP), A 27 KD IRON-SULFUR PROTEIN (IP), AND
CC TWO OTHER MEMBRANE-ANCHORING PROTEINS, QPS1 (CII-3) AND QPS2
CC (CII-4).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B560 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U57877; AAB41838.1; -.
CC EMBL; D49737; BAA31998.1; -.
CC EMBL; AF039594; AAC27993.1; -.
CC EMBL; AF039589; AAC27993.1; JOINED.
CC EMBL; AF039590; AAC27993.1; JOINED.
CC EMBL; AF039591; AAC27993.1; JOINED.
CC EMBL; AF039592; AAC27993.1; JOINED.
CC EMBL; AF039593; AAC27993.1; JOINED.
CC MIM; 602413; -.
CC PROSITE; PS01000; SDH_CVT_1; 1.
CC PROSITE; PS01001; SDH_CVT_2; 1.
CC PFAM; PF01127; sdh_cvt; 1.
CC Tricarboxylic acid cycle; Electron transport; Heme; Transmembrane;
CC Mitochondrion; Transit peptide.
CC TRANSIT 1 29
CC CHAIN 30 169
CC -----
CC TRANSMEM 71 93
CC TRANSMEM 105 127
CC TRANSMEM 147 168
CC BINDING 71 71
CC BINDING 127 127
CC SEQUENCE 169 AA; 18610 MW; 3F37EF06 CRC32;
CC -----
CC Query Match 14.7%; Score 82; DB 1; Length 169;
CC Best Local Similarity 28.0%; Pred. No. 7.22e+00;
CC Matches 14; Conservative 20; Mismatches 13; Indels 3; Gaps 3;
CC -----
CC Db 62 SLPMAWSICHTGTGALSAGVSL-FGMS-ALLLPFGNFES-YLELVKSLCL 108

Search completed: Mon Mar 20 12:31:26 2000
Job time : 9 secs.


```

01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SBF PROTEIN PRECURSOR.
GN SBF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
NC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RP SEQUENCE FROM N.A.
RA BOETTNER M., LAEFF M., SUTER-CRAZZOLARA C.;
RT "Identification of a novel member of the TGFbeta superfamily.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011969; CAA09891.1; -
DR EMBL; AJ011970; CAA09891.1; JOINED.
DR DR HSP; P18075; 1BMP.
FT SIGNAL.
KW FT SIGNAL.
SQ SEQUENCE 303 AA; 33438 MW; 01C35FCC CRC32;

Query Match 16.7%; Score 93; DB 11; Length 303;
Best Local Similarity 36.7%; Pred. No. 9.09e-01;
Matches 18; Conservative 11; Mismatches 18; Indels 2; G

Db 20 LFLLELLLSWPSQDALPERRSLSESNLPDELGRFDLLSRL 68
| : | | | | | : | : | : | : | : | : | : | : |
Qy 5 LPVLILLTLG-SHGTFGTQLK-LNESFLTSSVFLELLEKL 51
RESULT 3
ID Q17395 PRELIMINARY; PRT; 382 AA.
AC Q17395;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DI 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE APOCYTOCHROME B (EC 1.10.2.2) (UBIQUINOL-CYTOCHROME C REDUCTASE
DN COB.
GE Allomyces macrogynus.
CS Alomyces macrogynus.
OG Mitochondrion.
OC Eukaryota; Fungi; Chytridiomycota; Chytridiomycetes; Blastocladi
QC Blastocladiaceae; Allomyces.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE; 96226032.
RA PAQUIN B., LANG B.F.;
RT "The mitochondrial DNA of Allomyces macrogynus: the complete gen
RT RT sequence from an ancestral fungus.";
RJ J. Mol. Biol. 255:688-701(1996).
RN [2]
SEQUENCE FROM N.A.
RA PAQUIN B., LAFOREST M.J., LANG B.F.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C = Q + 2
CC FERROCYTOCHROME C.
DR EMBL; U41288; AAC49221.1; -.
DR PFAM; PF00032; cytochrome_b_c; 1.
DR PFAM; PF00033; cytochrome_b_n; 1.
KW Oxidoreductase; Mitochondrion.
SQ SEQUENCE 382 AA; 43467 MW; A2LEIABI CRC32;

Query Match 16.7%; Score 93; DB 8; Length 382;
Best Local Similarity 33.3%; Pred. No. 9.09e-01;
Matches 17; Conservative 15; Mismatches 16; Indels 3; G

Db 8 PVLSIANSFLIDPSFNNTYLNWNGSLG-LCLVIQIVGVYTLAHVAPS 57
| : | | | | | : | : | : | : | : | : | : | : |
QY 22 PGMTQLK-LKESFL-TNSVSYSFLELLEKLCLLLHPGSVTUHHARS 70
RESULT 4
ID O68567 PRELIMINARY; PRT; 209 AA.
AC O68567;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DI 01-AUG-1998 (TREMblrel. 07, Last sequence update)

```

```

AC Q9W6S7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE TRANSCRIPTION CO-REPRESSOR SIN3.
GN SIN3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RA VERMAAK D., WADE P.A., JONES P.L., SHI Y.-B., WOLFFE A.P.;
RT "Functional analysis of the SIN3-histone deacetylase: RPD3-RbAp48-
RT histone H4 connection in the Xenopus oocyte."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154112; AAB34644.1; -.
SQ SEQUENCE 1275 AA; 145088 MW; EB1942F4 CRC32;

Query Match 16.5%; Score 92; DB 13; Length 1275;
Best Local Similarity 27.5%; Pred. No. 1.22e+00;
Matches 11; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

Db 935 VGLKRDKNDSAIQRLKPEMDIAEDYNPAFDVMRL 974
QY 13 LLGSSHGTPGNTLQKLKESF-LTNSSYESSFLELLEKL 51

RESULT 7
ID P75377 PRELIMINARY; PRT; 879 AA.
AC P75377;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE FIL-ORF879 PROTEIN.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PIAGENS H., PIRKL E., LI B.C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=M129;
RA HIMMELREICH R., HILBERT H., LI B.-C.;
RL submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE000042; AAB96079.1; -.
SQ SEQUENCE 879 AA; 101086 MW; 9AD217E0 CRC32;

Query Match 15.9%; Score 89; DB 2; Length 879;
Best Local Similarity 28.9%; Pred. No. 2.92e+00;
Matches 11; Conservative 15; Mismatches 11; Indels 1; Gaps 1;

Db 677 EGIPKDSNY-SSFVHLDDKSLFLQAKVSGIDINENK 713
QY 32 ESFTNSSYESSFLELLEKLKLLHLPSTGTVLHAR 69

RESULT 8
ID O03305 PRELIMINARY; PRT; 297 AA.
AC O03305;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Carettochelys insculpta (pitted-shelled turtle).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ;
RA BOETTNER M., LAFF M., SUTER-CRAZZOLARA C.;

AC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Testudines;
OC Cryptodira; Trionychoidae; Carettochelyidae; Carettochelys.
RN [1]
RP SEQUENCE FROM N.A.
RA SHAFFER H.B., MEYLAN P., MCKNIGHT M.L.;
RL Syst. Biol. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C - Q + 2
CC -1- FERROCYTOCHROME C.
CC BOUND TO THE PROTEIN (BY SIMILARITY).
DR EMBL; U81355; AAB57646.1; -.
DR PFAM; PF00032; cytochrome_b_c; 1.
DR PFAM; PF00033; cytochrome_b_n; 1.
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
KW Heme.
FT NON_TER 297 297
SQ SEQUENCE 297 AA; 33587 MW; 7D5808B5 CRC32;

Query Match 15.8%; Score 88; DB 8; Length 297;
Best Local Similarity 31.5%; Pred. No. 3.89e+00;
Matches 17; Conservative 10; Mismatches 25; Indels 2; Gaps 2;

Db 165 GLTVHLLFLYETGSSNPTGLNSMD-KIPHPHYFSYKDFV-GLILMLAILNL 216
QY 4 GLPVLVLLTLGSSHGTPGNTLQKLKESF-LTNSSYESSFLELLEKLKLLNL 57

RESULT 9
ID Q69582 PRELIMINARY; PRT; 143 AA.
AC Q69582;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HERPESVIRUS TYPE 6 DNA.
OS Human herpesvirus 6.
OC Viruses; dsDNA Viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94181269.
RA THOMPSON J., CHOUDHURY S., KASHANCHI F., DONIGER J., BERNEMAN Z.,
RA FRENKEL N., ROSENTHAL L.J.;
RT "A transforming fragment within the direct repeat region of human
RT herpesvirus type 6 that transactivates HIV-1."
RL Oncogene 9:1167-1175(1994).
DR EMBL; X73675; CAA52028.1; -.
SQ SEQUENCE 143 AA; 13317 MW; 597857A6 CRC32;

Query Match 15.6%; Score 87; DB 14; Length 143;
Best Local Similarity 47.2%; Pred. No. 5.17e+00;
Matches 17; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

Db 88 LGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLGL 123
QY 1 MGSGLPLVLLTL-LGSSHGTPGNTLQKLKESFL 35

RESULT 10
ID Q92007 PRELIMINARY; PRT; 303 AA.
AC Q92007;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE SBF PROTEIN PRECURSOR.
GN SBF OR GDF15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ;
RA BOETTNER M., LAFF M., SUTER-CRAZZOLARA C.;

```

```

RT "Identification of a novel member of the TGF-beta superfamily.";
RN Submitter (OCT-1998) to the EMBL/GenBank/DBJ databases.
RX MEDLINE; 94150718.
RA HSTAO E.C., KONIARI L.G., ZIMMERS T.A., SEBALD S.M., SITZMANN J.V.,
RA HUYNH T.V., LEE S.-J.;
RT "Growth/differentiation factor-15: a new TGF-beta family member
RT induced following liver and bile duct injury.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011967; CAA09890.1; -.
DR EMBL; AJ011968; CAA09890.1; JOINED.
DR HSSP; F18075; 1BMP.
KW Signal.
FT SIGNAL
SQ SEQUENCE 303 AA; 33256 MW; FBBF45FE CRC32;

Query Match 15.6%; Score 87; DB 11; Length 303;
Best Local Similarity 36.7%; Pred. No. 5.17e+00;
Matches 18; Conservative 10; Mismatches 19; Indels 2; Gaps 2;

Db 20 LFLLLLLLLSWPSQGDALAMPQRRSGPESQNLNADLRGFODLLSRL 68
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5 LPLVLLTLTG-SHGTGPGNTLQKL-KESPLTNSSYESSFLELEKL 51
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
ID Q14968 PRELIMINARY; PRT; 322 AA.
AC Q14968;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE RHODOPSIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RX MEDLINE; 92250505.
RA KUNZ D., GERARD N.P., GERARD C.;
RT "The human leukocyte platelet-activating factor receptor. cDNA
RT cloning, cell surface expression, and construction of a novel epitope-
RT bearing analog.";
RL J. Biol. Chem. 267:9101-9106(1992).
DR EMBL; M76676; AAB97766.1; -.
DR PFAM; PF00001; 7tmL1; 2.
KW GTP-binding.
SQ SEQUENCE 322 AA; 33096 MW; 70F54EC8 CRC32;

Query Match 15.4%; Score 86; DB 4; Length 322;
Best Local Similarity 34.5%; Pred. No. 6.86e+00;
Matches 19; Conservative 13; Mismatches 22; Indels 1; Gaps 1;

Db 95 ALVLLIFLLSLGNCVGVVKKHQLRTVYNFLLSLSLDLTALCLPAA 148
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 6 PLVLLTLGSSHGTPGNTLQKLKESPLTNSSYESSFLELEKLCLLLHPSG 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
ID Q23220 PRELIMINARY; PRT; 562 AA.
AC Q23220;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE W08D2.3 PROTEIN.
GN W08D2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RX SEQUENCE FROM N.A.
RA SKINBURNE J., AINSCOUGH R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

```

```

RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z70271; CAA94232.1; -.
SQ SEQUENCE 562 AA; 63331 MW; 5272E400 CRC32;

Query Match 15.4%; Score 86; DB 5; Length 562;
Best Local Similarity 31.5%; Pred. No. 6.86e+00;
Matches 23; Conservative 13; Mismatches 27; Indels 4; Gaps 4;

Db 122 IILLICIVIVFGTGLGISLAVISN-FVIFK-QQSFL-LQQLHLIGAFSGVTAMFY 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 7 LVLLTLGSSHGTPGNTLQKLKESPLTNSSYESSFLELEKLCLLLHPSGTSVTLH 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 179 QSAPESEHOEDCNS 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 H-ARSOHHVVCNT 78
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
ID Q03295 PRELIMINARY; PRT; 297 AA.
AC Q03295;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Trachemys scripta (Red-eared slider turtle).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Testudines;
OC Cryptodira; Testudinoidae; Emydidae; Trachemys.
RN [1]
RX SEQUENCE FROM N.A.
RA SHAEFFER H.B., MEYLAN P., MCKNIGHT M.L.;
RL Syst. Biol. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C = Q + 2
CC -1- FERROCYTOCHROME C.
CC -1- COPACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
DR EMBL; U81351; AAB57666.1; -.
DR PFAM; PF00032; cytochrome_b_c; 1.
DR PFAM; PF00033; cytochrome_b_n; 1.
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
KW Heme.
FT NON_TER 1
FT NON_TER 297
SQ SEQUENCE 297 AA; 33479 MW; 1BCBCFAC CRC32;

Query Match 15.2%; Score 85; DB 8; Length 297;
Best Local Similarity 33.3%; Pred. No. 9.08e+00;
Matches 18; Conservative 9; Mismatches 25; Indels 2; Gaps 2;

Db 165 GLTLVHLLFLHETGNSNPTGLNSNVD-KIPFHPYFSYKD-LLGIILMLTLTL 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4 GLPLVLLTLGSSHGTPGNTLQKLKESPLTNSSYESSFLELEKLCLLLH 57
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
ID Q03294 PRELIMINARY; PRT; 297 AA.
AC Q03294;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)

```

Wed Mar 22 12:22:52 2000

[illegible]

Search completed: Mon Mar 20 12:32:03 2000
Job time : 19 secs.

